

VISIT...

LANZAROTE
Caliente.COM

Lecture, Week 1: September 27th - October 3rd, 1999

Outline

1. What is consciousness? An ancient debate
2. The study of consciousness in the 20th century
3. The prevalent view: computational emergence
4. The quantum approach to consciousness
5. *Conclusion*: Is consciousness due to computational emergence or fundamental quantum process?

1. What is consciousness? An ancient debate

How does brain activity produce conscious experience? Why do we feel love, hear a flute, see a purple sunset? Philosophers call the raw components which comprise conscious experience qualia (e.g. Chalmers, 1996).

It is not at all obvious why we need qualia from an evolutionary standpoint---complex, adaptive behavior of unfeeling zombie-like creatures might well have enabled them to flourish. However it seems unlikely that, for example, a zombie could have painted the Mona Lisa. So what are these raw feelings? What is consciousness?

Since ancient Greece there have been two types of explanations. On the one hand Socrates said that conscious mental phenomena were "products of the cerebrum"---produced by activities in the brain. On the other hand Democritus believed mental phenomena were fundamental "features of reality", merely accessed by the brain.

Figure 1. Are mental 'qualia' like the redness of a rose, patterns of activity within the brain, or a fundamental feature of nature?

2. The study of consciousness in the 20th century

In the past hundred years scientific pursuit of consciousness has had its ups and downs. William James's 1903 "Principles of Psychology" had placed the topic at center stage, but then for most of the century behaviorist psychology relegated consciousness to a role as irrelevant epiphenomenon. Neurophysiologists kept the term consciousness alive during these "Dark Ages", and in the 1960's and 1970's cognitive functions became correlated with the brain's inner states. Nevertheless the "C-word" remained slightly off-color: "Why discuss something that can't be measured?" In the 1990's much of this changed and the problem of consciousness became a hot topic spanning many fields. As the millennium nears, consciousness is coming to the forefront of scientific frontiers.

There are numerous reasons for the late 20th century consciousness boom, but credit justifiably goes to scholarly and passionate books written by luminaries such as Roger Penrose (1989; 1994), Francis Crick (1994), John Eccles (1989) and Gerald Edelman (1989) who each digressed from primary fields to address their intimate mystery. These noble efforts were preceded by Marcel and Bisiach's anthology *Consciousness in*

Contemporary Science (1988) and Bernard Baar's A Cognitive Theory of Consciousness (1988).

Within philosophy a focus of disagreement regarding the nature of conscious experience developed in the 1970's and 1980's. "Anti-physicalists" questioned how conscious experience could be derived from ordinary matter (a view derived from Socrates and espoused in slightly varying forms by

"reductionists/functionalists/materialists/physicalists"). The key figures producing this anti-physicalist pressure were Saul Kripke (1972), Thomas Nagel (1974) and Frank Jackson (1982). On the other side Patricia Churchland (1986), Paul Churchland (1989) and others steadied the flag of physicalism, and Daniel Dennett's flamboyantly titled Consciousness Explained (1991) pushed it overboard by arguing conscious experience out of existence. However David Chalmers' The Conscious Mind (1996) illuminated the chasm between physicalist explanations and the facts about consciousness (the "hard problem"), and suggested conscious experience may be an irreducible, fundamental property of reality (Rosenberg, 1997).

3. The prevalent view: computational emergence

Despite some anti-physicalist pressure, the modern version of Socrates' view is the currently dominant scientific position. Conventional explanations portray consciousness as an emergent property of computer-like activities in the brain's neural networks. The brain is essentially a computer, and neuronal excitations (axonal firings/synaptic signaling) are fundamental information states, or "bits" equivalent to either 1 or 0. The prevailing views among scientists in this camp are that 1) patterns of neural network activities correlate with mental states, 2) synchronous network oscillations in thalamus and cerebral cortex temporally bind information, and 3) consciousness emerges as a novel property of computational complexity among neurons.

Figure 2. PET scan image of brain showing visual and auditory recognition (from S Petersen, Neuroimaging Laboratory, Washington University, St. Louis. Also see J.A. Hobson "Consciousness," Scientific American Library, 1999, p. 65).

Yet computers aren't conscious, at least as far as we can tell. To explain how qualia and consciousness may be produced by complex interactions among simple neurons, proponents of emergence theory point out that it is quite common in nature for complex interactions among many simple states to lead to higher order emergence of phases with novel properties. Examples include a candle flame, whirlpool, or the Great Red Spot on Jupiter. Could conscious experience be such an emergent, novel property? (See Alwyn Scott's "Stairway to the Mind" for an elegant exposition of the emergence argument).

Figure 3. Electrophysiological correlates of consciousness.

But weather patterns and other known emergent phenomena aren't conscious either. Complex neuronal interactions are indeed essential to our form of consciousness, and do appear to correlate with conscious experience. But at what critical level of computational complexity does consciousness emerge? Why do the emergent properties have qualia? Is some purely biological factor necessary for the emergence of conscious experience? Or are deeper level factors in play?

For a debate between Alwyn Scott and Stuart Hameroff on the merits of emergence theory vs quantum approaches to explain consciousness, see the article "Sonoran afternoon" in the Tucson II book, and at

[~http://www.u.arizona.edu/~hameroff/sonoran.html](http://www.u.arizona.edu/~hameroff/sonoran.html)

[<http://www.u.arizona.edu/hameroff/sonoran.html>](http://www.u.arizona.edu/hameroff/sonoran.html)

In addition to 'qualia' computational emergence approaches appear to fall short in fully explaining other enigmatic features of consciousness:

- The nature of subjective experience, or 'qualia'- our 'inner life' (Chalmers' "hard problem");
- Subjective binding of spatially distributed brain activities into unitary objects in vision, and a coherent sense of self, or 'oneness';
- Transition from pre-conscious processes to consciousness itself;
- Non-computability, or the notion that consciousness involves a factor which is neither random, nor algorithmic, and that consciousness cannot be simulated (Penrose, 1989, 1994, 1997);
- Free will;
- Subjective time flow.
- Apparent reverse time anomalies (e.g. Libet, Radin/Bierman)

As we shall see, these enigmatic features may be amenable to quantum approaches.

Another problem with computational emergence theory is that in fitting the brain to a computational view, such explanations omit incompatible neurophysiological details:

- Widespread apparent randomness at all levels of neural processes (is it really noise, or underlying levels of complexity?);
- Glial cells (which account for some 80% of brain);
- Dendritic-dendritic processing;
- Electrotonic gap junctions;
- Cytoplasmic/cytoskeletal activities; and,
- Living state (the brain is alive!).

A further difficulty is the absence of testable hypotheses in emergence theory. No threshold or rationale is specified; rather, consciousness "just happens".

Finally, the complexity of individual neurons and synapses is not accounted for in such arguments. Since many forms of motile single-celled organisms lacking neurons or synapses are able to swim, find food, learn, and multiply through the use of their internal cytoskeleton, can they be considered more advanced than neurons?

Figure 4. Single cell paramecium can swim and avoid obstacles using its cytoskeleton. It has no synapses. Are neurons merely simple switches?

4. The quantum approach to consciousness

An alternative to Socrates' view and that of current emergence theory stems from Democritus who believed qualia to be primitive, fundamental aspects of reality, irreducible to anything else. Philosophical renditions along these lines have included panpsychism (e.g. Spinoza, 1677), panexperientialism (e.g. Whitehead, 1920) and most recently pan-protopsyhism (Chalmers, 1995). Perhaps most compatible with modern physics is Whitehead who believed consciousness to be a sequence of discrete events ("occasions of experience") occurring in a wider, proto-conscious field. (Abner Shimony in 1993 pointed out that Whitehead's discrete events were consistent with quantum state reductions.) (Lecture week 3)

Could Whitehead's "philosophical" proto-conscious field be the basic level of physical reality? Do qualia exist as fundamental features of the universe (e.g. like spin or charge), somehow accessed by brain processes to adorn neuronal activities with conscious experience?

What physical features of the universe could relate to qualia? Can qualia be given a physical correlate, say in modern descriptions of the fundamental nature of the universe? Whether or not the fundamental nature of reality---empty space---is a true void or has some underlying structure is a question which also dates to the Greeks. Democritus saw a true void, while Aristotle saw a background pattern or "plenum" with 3 dimensions. In the 17th century Galileo described the motion of free-falling objects, and Newton discovered that the force of gravity between two objects depends on the mass of the objects and the distance between them. But Newton couldn't understand how such a mysterious force could operate in the absolute void he assumed empty space to be. In the 19th century Maxwell postulated a "luminiferous ether" as a background pattern to explain the propagation of electromagnetic waves in a vacuum, but the Michelson-Morley experiment seemed to rule out a background medium. Einstein's special relativity seemed to confirm the view of a true void with its lack of preferred reference frames, although it did introduce the concept of a four dimensional universe in which the 3 dimensions of space are unified with the dimension of time.

Einstein's general relativity stated that a massive object such as the sun curves spacetime around itself, much as a bowlingball would form a depression in a rubber sheet. Smaller objects such as the earth and other planets move around the object like marbles would roll down the depression in the rubber sheet. Therefore gravity was not a mysterious force but curvatures in reality itself, what Einstein called the spacetime metric.

Figure 5. Einstein's spacetime metric describes the curvature of spacetime at large scales (Science Year 2000).

Einstein's general relativity with its curvatures and spacetime metric swung the prevalent scientific view back to an underlying pattern. But what exactly is the spacetime metric? And what is happening at the smallest scales? We do know that at the level of the Planck

scale (10^{-33} cm, 10^{-43} sec) spacetime is no longer smooth, but quantized. The nature of reality at the level of atoms, sub-atomic particles and below remained mysterious. These effects are approached through quantum mechanics, a branch of physics developed in the early 1900's by Niels Bohr of Denmark, Erwin Schrodinger of Austria, and Werner Heisenberg of Germany. Quantum mechanics explains how atoms absorb and give off units of energy called quanta. Surprisingly, quanta act as both particles and waves, and can exist in quantum superposition of two or more states or locations simultaneously! While quantum mechanics cannot explain how this is possible, it can explain the behavior of the various particles that make up atoms.

Physicists have developed a quantum theory known as the Standard Model of Particle Physics to describe all of the fundamental particles that make up atoms. Protons and neutrons which make up the atom's nucleus are themselves made up of tiny particles called quarks. There are 6 kinds of quarks, and atoms also contain electrons that orbit the nucleus. The Standard Model also describes the forces at work at the sub-atomic realm, including the electromagnetic force (which holds atoms and molecules together), the strong force (which holds protons and neutrons together), and the weak force (which causes certain types of radioactive decay). The forces are carried by particle/waves called bosons, which include photons which carry the electromagnetic force. The Standard Model predicts another as-yet-unobserved particle/wave, the Higgs boson which particle physicists believe may give mass to particles. The Standard Model does not include a description of gravity, which in the atomic realm is much weaker than the other forces. Another odd feature of quantum particle/waves is quantum entanglement. If two quantum particles are coupled but then go their separate ways, they remain somehow connected over space and time. Measurement of one will affect the state of the other (Lecture Week 2). Also, quantum particles can condense into one collective quantum object in which all components are governed by the same wave function (e.g. Bose-Einstein condensate). But we don't see this quantum weirdness in our seemingly classical world. Science has one theory (general relativity) to describe the behavior of large masses such as planets and pencils, and another theory (quantum mechanics) to explain the behavior of atoms and sub-atomic particles. What is the cut-off between the quantum and classical worlds? What is needed is a "unified theory" combining the large scale and the small scale. One approach to a unified theory is found in Roger Penrose's view of the meaning of quantum superposition. Quantum theory tells us that at small scales particles may have no definite location or state, and exist in "superposition" of many possible states ("wave-particle duality"). What does it mean to be in two or more states or locations simultaneously? Could this property be relevant, or useful in understanding consciousness? It turns out quantum superposition is indeed relevant to a new form of technological computing - quantum computing. Could this property be relevant, or useful in understanding consciousness? It turns out quantum superposition is indeed relevant to a new form of technological computing - quantum computing.

Figure 6. Wave-particle duality of quantum objects is essential to quantum computing. A particle has wave-like quantum behavior, and may also localize to a definite state/location. From "Explorations in Quantum Computing", by Colin P. Williams and Scott H. Clearwater, Springer-Verlag, New York, 1998

Quantum computing is a proposed technology which aims to take significant advantage of features of quantum theory, and may connect brain activity to fundamental spacetime geometry.

Described theoretically in the 1980's (e.g. by Benioff, Feynman, Deutsch), "quantum computing" is suggested to utilize quantum superposition, in which a particle can exist in two or more states, or locations simultaneously. Whereas current computers represent information as "bits" of either 1 or 0, quantum computers are proposed to utilize quantum bits---"qubits"---of both 1 AND 0.

Potential advantages in quantum computing stem from qubits in superposition interacting nonlocally by quantum coherence or entanglement, implementing near-infinite quantum parallelism. These interactions perform computations and, at some point, the qubits all collapse, or reduce to a single set of classical bits---the "solution".

Significant technological and commercial advantages are offered by quantum computers if they are ever actually constructed (prototypes now exist, and research activity is intense). But regardless, theoretical aspects of quantum computing teach an enormous amount about fundamental physics, and possibly consciousness.

The essential feature in quantum computing is superposition---each qubit may be in superposition of both 1 and 0. What exactly is superposition? How are we to understand this basic feature of unobserved quantum particles? How can something be in two places or two states at once?

Roger Penrose explains superposition in the following way. According to Einstein's general relativity, mass is equivalent to curvature in fundamental spacetime geometry.

The fundamental feature is not mass, but spacetime geometry. Mass existing in one particular location or state is ("in reality") one particular curvature in fundamental spacetime geometry.

Figure 7. According to Einstein's general relativity, mass is equivalent to curvature in spacetime geometry. Penrose applies this equivalence to the fundamental Planck scale. The motion of an object between two conformational states of a protein such as tubulin (top) is equivalent to two curvatures in spacetime geometry as represented as a two-dimensional spacetime sheet (bottom).

Now consider quantum superposition. In Penrose's view superposition of a particle in two states or locations implies superposition of two different spacetime curvatures---an unstable separation, or bubble in spacetime itself.

Figure 8. Mass superposition, e.g. a protein occupying two different conformational states simultaneously (top) is equivalent, according to Penrose, to simultaneous spacetime curvature in opposite directions - a separation, or bubble ("blister") in fundamental spacetime geometry. Quantum computation would therefore be an organized process of separations and reductions at the fundamental level---reconfiguring the basic level of spacetime geometry. If qualia are embedded at this same basic level, then some form of quantum computation in the brain could access fundamental qualia and select particular patterns of conscious experience.

Figure 9. Are mental 'qualia' like the redness of a rose fundamental patterns in spacetime geometry? This connection between the brain and fundamental spacetime geometry could patch the age-old controversy. Socrates and Democritus were both right---consciousness requires brain processes which access and select fundamental qualia embedded at the basic level of spacetime geometry. Our conscious experience could derive from some basic process rippling through reality.

5. Conclusion: Is consciousness due to computational emergence or fundamental quantum process?

The burning issue is thus whether or not conscious experience like feelings, qualia, our "inner life" can be accommodated within present-day science. Those who believe it can (e.g. physicalists, reductionists, materialists, functionalists, computationalists, emergentists) - like Socrates - see conscious experience as an emergent property of complex neural network computation. Others see conscious experience either outside science (dualists), or like Democritus believe science must expand to include conscious experience as a primitive feature of nature, something like spin or charge (idealists, panpsychists, pan-experientialists, pan-protopsyhists).

Here is how (from an admittedly biased perspective) computational emergence views stack up against quantum approaches in attempting to explain the enigmatic features of consciousness.

Feature 1: 'Qualia', experience, hard problem (e.g. Chalmers, 1996a; 1996b)

Computational emergence: Postulate conscious experience emerging from critical level of neuronal computational complexity.

Unanswered: What is the critical level? Why should emergent phenomenon have conscious experience? What causes transition from pre-conscious processes to consciousness itself?

Quantum approaches: May philosophically subscribe to panexperientialism, or pan-protopsyhism following e.g. Spinoza, Leibniz, Whitehead, Wheeler and Chalmers. Experience, qualia are deemed fundamental, a basic property of the universe like spin or charge, perhaps rooted at the Planck scale.

Unanswered: What are fundamental properties? What unifying principle exists (we don't want a new fundamental property for every conceivable experience).

How can the brain link to the quantum level or Planck scale?

Feature 2: Binding, unity in vision and imagery, sense of self, 'oneness'

Computational emergence: Coherent neural membrane activities (e.g. coherent 40 Hz) bind by temporal correlation of neural firing.

Unanswered: Why should temporal correlation bind qualia when different features processed asynchronously are nonetheless somehow consciously experienced as synchronous? (Zeki, 1998).

Quantum approaches: Macroscopic quantum states (e.g. Bose-Einstein condensates) are single entities, not just temporally correlated. Unity is provided through non-locality - components are intimately connected and governed by a single wave function.

Unanswered: How can macroscopic quantum states exist in the brain and interact with neural structures?

Feature 3. Transition from pre-conscious processes to consciousness itself

Computational emergence: Conscious experience emerges at a critical threshold of neural activity?

Unanswered: What is the critical level? Why is the emergent phenomenon "conscious"? What is the essential difference between non-conscious, or pre-conscious neural processes, and conscious neural processes?

Quantum approaches: Pre-conscious processes are in quantum superposition of possible states. The process by which the possibilities reduce/collapse to definite states is the transition between pre-conscious and conscious processes.(e.g. Stapp, Penrose-Hameroff)

Unanswered: How is the quantum superposition manifest in the brain? What is the cause of collapse? How is the output implemented by neural structures?

Feature 4: Non-computability (Penrose argues from Godel's theorem that human thought has a non-computable component - some feature which is neither algorithmic nor random/probabilistic).

Computational emergence: What non-computability?

Unanswered: How do we differ from computers?

Quantum approaches: Penrose quantum gravity collapse (objective reduction) of the wave function is proposed to be non-computable (influenced by Platonic influences ingrained in spacetime geometry).

Unanswered: How does objective reduction link to the brain?

Feature 5: Free will (The problem in understanding free will is that our actions seem neither deterministic nor random (probabilistic). What else is there in nature?

Computational emergence: Doesn't offer much help. Processes are either deterministic or random. Questions existence of free will.

Unanswered: Everything.

Quantum approaches: Penrose's objective reduction ("OR") is a possible solution. Conscious actions are products of deterministic processes acted on at moments of reduction by non-computable 'Platonic' influences intrinsic to fundamental spacetime geometry.

Unanswered: How are Platonic influences represented? How do they influence neuronal structures?

Feature 6: Subjective time flow Physics has no requirement for a forward flow of time, yet our conscious experience seems to have a flow of time.

Computational emergence: Agnostic - no position.

Unanswered: Why does time seem to flow?

Quantum approaches: Quantum state reductions/collapses are irreversible. A series of quantum reductions would "ratchet" through time to give an experience of time flow.

Unanswered: Where, how in the brain could quantum state reductions occur?

Feature 7. Reverse time anomalies. Research by Ben Libet in the late 70's and recent work by Dean Radin and Dick Bierman suggest that somehow the brain can refer information (or qualia) backwards in time.

Computational emergence: No position - refute/explain away data.

Unanswered: What's wrong with the data interpretation leading to conclusions regarding backward time referral? If experimentally corroborated, how can backwards time referral be explained?

Quantum approaches: Intervals between quantum state reductions are 'atemporal', and reductions may send quantum information backwards through time.

Unanswered: How do reductions occur? How is quantum information processed? Are qualia equivalent to quantum information?

So in an admittedly biased survey, quantum approaches are far better equipped to deal with the enigmatic features of consciousness than are conventional computationalist emergent approaches.

It is probably inevitable that the brain/mind be compared to a quantum computer. The brain/mind has been historically compared to the current most advanced form of technological information processing. In ancient Greece the mind was compared to a seal ring in wax', in the 19th century a telegraph switching circuit, and in the 20th century a hologram, and presently to the most deeply developed and successful metaphor - the classical computer. If quantum computers become technological reality and supercede classical computers as our most advanced form of information processing technology, the brain/mind will be expected to be at least equally advanced, and hence utilize some type of quantum computation. Even if quantum computers do not become successful technology, their mere theoretical existence has profound implications for fundamental physics, and consciousness.

Outline

Pan-protopsychism

The fundamental nature of reality

Is there a connection between the brain and "funda-mental" reality?

Illustrations

Penrose's Three Worlds (after Popper) *3worlds.jpg*

The 'quantum foam' (from Kip Thorne) *foam.jpg*

The 'Casimir force' *casimir.jpg*

String theory *string.gif*

Violin analogy for string theory *violin.jpg*

Quantum geometry portrays spacetime as a woven fabric *fabric.jpg*

Planck scale spin networks may be the basic level of reality *spinnet.jpg*

Mass equivalent to spacetime curvature *spacecurv.jpg*

Superposition is equivalent to separation (bubble, or blister) in spacetime geometry *spacecurv1.jpg*

Classical and superpositioned spacetime geometry *spacecurv2.jpg*

Classical and superpositioned mass/spacetime form 'qubits' useful in quantum computation.

spacecurv3.jpg

1. Pan-protopsychism

If computational emergence is unable to account for conscious experience and other enigmatic features of consciousness, what approach can do so?

A line of panpsychist, panexperiential philosophy stemming from Democritus suggests that proto-conscious experience is fundamental. In the past 100 years quantum theory and modern physics have examined the basic nature of reality, and at the millenium these two avenues may be converging.

An extreme panpsychist view is that consciousness is a quality of all matter: atoms and their subatomic components having elements of consciousness (e.g. Spinoza, 1677; Rensch, 1960). "Mentalists" such as Leibniz and Whitehead (e.g. 1929) refined this position, contending that systems ordinarily considered to be physical are constructed in some sense from mental entities. Leibniz (e.g. 1768) saw the universe as an infinite number of fundamental units ("monads") each having a primitive psychological being. Whitehead (e.g. 1929) described dynamic monads with greater spontaneity and creativity, interpreting them as mind-like entities of limited duration ("occasions of experience" each bearing a quality akin to "feeling").

Bertrand Russell (1954) described "neutral monism" in which a common underlying entity, neither physical nor mental, gave rise to both. More recently Wheeler (e.g. 1990) described a "pre-geometry" of fundamental reality comprised of information.

Chalmers (1996a;1996b) contends that fundamental information includes "experiential aspects" leading to consciousness. Chalmers coined the term "pan-protopsychism" to allow for an interaction between the brain and fundamental entities producing consciousness as we know it. Consciousness could be a product of the brain's interaction with fundamental reality.

Idealism is the notion that "consciousness is all there is". That is, our conscious minds create the external world. A problem for idealism is intersubjective agreement, how elements of reality usually thought of as 'external' or 'physical' are common to any conscious being. This is also known as the problem of "solipsism", which occurs in approaches in which the self is the total of existence. Universal mind may address this concern, as agreement might be arrived at through sharing thoughts rather than a physical reality. The challenge is then to explain why some elements are shared and others are absolutely private. In Eastern traditions, universal "Mind" is often seen as fundamental, accessible to the brain (e.g. Goswami, 1993). And some Buddhist meditative disciplines portray consciousness as sequences of individual, discrete events; trained meditators describe distinct "flickerings" in their experience of reality (Tart, 1995). Buddhist texts portray consciousness as "momentary collections of mental phenomena", and as "distinct, unconnected and impermanent moments which perish as soon as they arise." Each conscious moment successively becomes, exists, and disappears - its existence is instantaneous, with no duration in time, as a point has no length. Our normal perceptions, of course, are seemingly continuous, presumably as we perceive "movies" as continuous despite their actual makeup being a series of frames. Some Buddhist writings even quantify the frequency of conscious moments. For example the Sarvaastivaadins (von Rospatt, 1995) described 6,480,000 "moments" in 24 hours (an average of one "moment" per 13.3 msec), and some Chinese Buddhism as one "thought" per 20 msec. Perhaps not surprisingly, these times

correspond nicely with neurophysiological events such as coherent 40 Hz (~ 25 msec), and are consistent with proposals for quantum state reductions in the brain (as we shall see in later lectures).

This suggests not only that consciousness is a sequence of discrete events (experienced continuously, perhaps in a way like a movie is perceived continuously despite being comprised of discrete frames), but also that consciousness is serial! Recent evidence suggests that visual consciousness is indeed serial. See:

<http://www.eetimes.com/story/technology/advanced/OEG19990908S0001>

If qualia are embedded in fundamental reality accessible to the brain, perhaps Platonic values may be there as well. Plato in 400 B.C.: "...ideas have an existence...an ideal Platonic world...accessible by the intellect only...a direct route to truth...". In "Shadows of the Mind" Penrose (1994) described three worlds: the physical world, the mental world and the Platonic world. The physical world and the mental world are familiar and agreed upon as actual realities clearly, the physical world exists and thoughts exist.

Penrose's Platonic world includes mathematical truths, laws and relationships, as well as primitives for aesthetics and ethics---affecting our senses of beauty and morality. The Platonic world appears purely abstract. Could it simply exist in the empty space of the universe? If essential aspects of truth and beauty are indeed fundamental, perhaps the Platonic world is ingrained at the most basic level of reality along with qualia (perhaps qualia and Platonic values are the same?).

A problem for realistic Platonism is that concepts like truth and beauty would seem to be culturally-loaded, complex abstractions constructed according to experience and built up from a vast multitude of simpler generalizations and categorizations. But perhaps what is embedded in fundamental reality are primitive elements of these complex abstractions, distributed nonlocally through spacetime geometry.

Another objection is that Platonic values should be experienced in the same way by everyone. But we don't all like the same food, or favorite movie star. However Platonic values would be interacting with varying cultural, genetic, memory and other influences to give a final response. Plus, fundamental patterns in spacetime geometry and Platonic values embedded there may be evolving over time (as suggested by Lee Smolin in "Life of the Cosmos").

Figure 1. Penrose's 3 worlds (after Popper) "In some way, each of the three worlds, the Platonic mathematical, the physical, and the mental, seems to mysteriously emerge from---or at least be related to---a small part of its predecessor (the worlds being taken cyclically)." Shadows of the Mind. But could qualia and Platonic values be embedded in spacetime geometry, in empty space? What is empty space?

2. The fundamental nature of reality

As described briefly in Week 1, the nature of space and time has also been debated since the ancient Greeks. Democritus (who thought 'atoms' were the basic constituents of matter) saw space as a true, empty void through which matter traveled. On the other hand Aristotle saw empty space containing a background pattern, which he termed the 'plenum'. In the 19th century consideration of how electromagnetic waves travel through a vacuum led Maxwell to propose his "luminiferous ether", however the famous Michelson-Morley experiment seemed to refute such an idea and support again the notion of a true void. Einstein's special relativity with no preferred frames of reference also came down on the side of a true void. However Einstein's general relativity with spacetime curvature weighed in on the side of a background pattern in spacetime which Einstein termed the 'metric'. Since then numerous ideas and theories have attempted to describe some makeup of fundamental spacetime geometry.

A branch of quantum field theory known as quantum electrodynamics (QED) predicts that virtual particles and waves (virtual photons, matter-antimatter pairs) continuously

wink into and out of existence (e.g. Jibu and Yasue, 1995; Seife, 1997). These are quantum fluctuations that impart dynamic structure to the vacuum. The energy of the virtual particles contributes to the zero-point energy of the vacuum. At a much higher-energy (or equivalently, much smaller in scale) level, fluctuations in the topology of spacetime lead to what is described as the "quantum foam".

Figure 2. The "quantum foam": Quantum fluctuations in the topology of spacetime produce a foam of erupting and collapsing superpositions of spacetime topologies. (from Thorne, 1994)

This picture of the quantum vacuum had been developed by Max Planck and Werner Heisenberg in the 1920's. In 1948 the Dutch scientist Hendrick Casimir predicted that the all-pervading zero point energy could be measured using parallel surfaces separated by a tiny gap. Some (longer wavelength) virtual photons would be excluded from the gap region, Casimir reasoned, and the surplus photons outside the gap would exert pressure forcing the surfaces together. Recently, this "Casimir force" was quantitatively verified quite precisely (Lamoreaux, 1997), confirming the zero point energy. Lamoreaux's experimental surfaces were separated by a distance d ranging from 0.6 to 6 microns, and the measured force was extremely weak (Figure 3a).

Figure 3. A: The Casimir force of the quantum vacuum zero point fluctuation energy may be measured by placing two macroscopic surfaces separated by a small gap d . As some virtual photons are excluded in the gap, the net "quantum foam" exerts pressure, forcing the surfaces together. In Lamoreaux's (1997) experiment, d was in the range 0.6 to 6.0 microns (~1500 nanometers). B: Hall (1996; 1997) calculated the Casimir force on microtubules. As the force is proportional to d^{-4} , and d for microtubules is 15 nanometers, the predicted Casimir force is roughly 106 greater on microtubules (per equivalent surface area) than that measured by Lamoreaux. Hall calculates a range of Casimir forces on microtubules (length dependent) from 0.5 to 20 atmospheres.

Could the Casimir force have biological influence? At the Tucson II conference, physicist George Hall (Hall, 1996; 1997) presented calculations of the Casimir force on model cylinders representing biological microtubules which form networks within neurons and other cells. Hall considered the microtubule hollow inner core of 15 nanometers diameter as the Casimir gap d . (Hall assumed microtubules are electroconductive; recent experiments have shown this to be true). As the force is predicted to be proportional to d^{-4} , Hall's models predict significant pressure (0.5 to 20 atmospheres) exerted by the quantum vacuum on microtubules of sufficient length (Figure 3b). Microtubules actually are under compression in cells, a factor thought to enhance vibrational signaling and tensegrity structure (e.g. Ingber, 1993). In the well known phenomenon of "pressure reversal of anesthesia," unconscious, anesthetized experimental subjects wake up when ambient pressure is increased on the order of 10 to

100 atmospheres. This implies that a baseline ambient pressure such as the Casimir force acting on microtubules as suggested by Hall may be required for consciousness.

Despite these measurable effects, the fundamental nature of spacetime and the unification of general relativity and quantum mechanics remain elusive. However several approaches to the deepest structure of spacetime have emerged. What is known is that the Planck scale (10^{-33} cm, 10^{-43} sec) is the scale at which spacetime is no longer smooth, but quantized.

One approach is string theory, or superstrings, begun by John Schwarz of CalTech and Michael Green of Queen Mary College in London in the 1980's. The theory combines general relativity and quantum mechanics, and also can explain exotic entities which comprise all particles in the universe.

String theory states that quarks, electrons and other particles of matter, rather than being point-like, are actually tiny "line-like" objects called strings. The strings are incredibly small---each is approximately the size of the Planck length (10^{-33} cm), the smallest possible distance in spacetime, one billionth of one billionth the radius of an electron. Just as the strings of a violin can vibrate in many ways to produce varying musical notes, the tiny strings in spacetime can vibrate in many ways to create the different types of elementary particles. Strings also potentially explain the force-carrying particle/waves (e.g. bosons) which act on matter. Further, as the strings move within curved spacetime, string theory may unify general relativity and quantum mechanics.

Figure 4. String theory predicts matter and forces arise from vibrating Planck scale strings (Science Year 2000)

Figure 4a. Vibrations of fundamental strings are proposed to result in particles and energy, like the vibrations of violin strings produce sound.

String theory predicts that the strings can exist in two basic forms---open and closed. Strings interact by splitting and joining, and with vibratory motion can thus determine the particles and forces which make up the universe.

The problems with string theory are that 1) it is currently untestable, and 2) it requires 10 or 11 dimensions. Do these extra dimensions actually exist, or are they mathematical abstractions with no basis in reality?

Another approach to the fundamental nature of reality which requires only 4 dimensional spacetime is quantum geometry (Figure 5). To provide a description of the quantum mechanical geometry of space at the Planck scale, Penrose (1971) introduced "quantum spin networks" in which spectra of discrete Planck scale volumes and configurations are obtained (Figure 6). Ashtekar, Rovelli and Smolin have generalized these spin networks to describe quantum gravity. These fundamental spacetime volumes and configurations

may qualify as philosophical (quantum) monads. Perhaps quantum gravity Planck-scale spin networks encode proto-conscious experience and Platonic values?

Figure 5. Quantum geometry - the tiniest units of space consist of a complex fabric of interwoven threads which give rise to the spacetime continuum (Science Year 2000)

Figure 6. A quantum spin network. Introduced by Roger Penrose (1971) as a quantum mechanical description of the geometry of space, spin networks describe spectra of discrete Planck scale volumes and configurations (with permission, Rovelli and Smolin, 1995).

There are reasons to suspect gravity, and in particular quantum gravity in the fundamental makeup of spacetime geometry.

Roger Penrose: "The physical phenomenon of gravity, described to a high degree of accuracy by Isaac Newton's mathematics in 1687, has played a key role in scientific understanding. However, in 1915, Einstein created a major revolution in our scientific world-view. According to Einstein's theory, gravity plays a unique role in physics for several reasons (cf. Penrose, 1994). Most particularly, these are: 1) Gravity is the only physical quality which influences causal relationships between space-time events. 2) Gravitational force has no local reality, as it can be eliminated by a change in space-time coordinates; instead, gravitational tidal effects provide a curvature for the very space-time in which all other particles and forces are contained.

It follows from this that gravity cannot be regarded as some kind of "emergent phenomenon," secondary to other physical effects, but is a "fundamental component" of physical reality. It may seem surprising that quantum gravity effects could plausibly have relevance at the physical scales relevant to brain processes. For quantum gravity is normally viewed as having only absurdly tiny influences at ordinary dimensions. However, we shall show later that this is not the case, and the scales determined by basic quantum gravity principles are indeed those that are relevant for conscious brain processes.

Perhaps we may view the elusive quantum gravity as the basic makeup of fundamental spacetime geometry. But how is it connected to brain processes?

3. Is there a connection between the brain and 'fundamental' reality?

In David Bohm's approach, fundamental qualia might play the role of the quantum potential. Bohm's approach will be discussed in Week 5 by Paavo Pylkkanen.

Another approach is that of Michael Conrad's, called the fluctuon model. This is summarized in a paper accompanying this lecture.

In the approach taken by Roger Penrose, Einstein's general relativity is combined with aspects of quantum theory to give a process (Objective Reduction:OR) occurring at the fundamental level of spacetime geometry.

Penrose: (from "Conscious events as orchestrated spacetime selections")

~<http://www.u.arizona.edu/~hameroff/hardfina.html>

<<http://www.u.arizona.edu/hameroff/hardfina.html>>

According to modern accepted physical pictures, reality is rooted in 3-dimensional space and a 1-dimensional time, combined together into a 4-dimensional space-time. This space-time is slightly curved, in accordance with Einstein's general theory of relativity, in a way which encodes the gravitational fields of all distributions of mass density. Each mass density effects a space-time curvature, albeit tiny.

This is the standard picture according to classical physics. On the other hand, when quantum systems have been considered by physicists, this mass-induced tiny curvature in the structure of space-time has been almost invariably ignored, gravitational effects having been assumed to be totally insignificant for normal problems in which quantum theory is important. Surprising as it may seem, however, such tiny differences in space-time structure can have large effects, for they entail subtle but fundamental influences on the very rules of quantum mechanics.

Figure 7. According to Einstein's general relativity, mass is equivalent to curvature in spacetime geometry. Penrose applies this equivalence to the fundamental Planck scale. The motion of an object between two conformational states of a protein such as tubulin (top) is equivalent to two curvatures in spacetime geometry as represented as a two-dimensional spacetime sheet (bottom).

Figure 8. Mass superposition, e.g. a protein occupying two different conformational states simultaneously (top) is equivalent, according to Penrose, to simultaneous spacetime curvature in opposite directions - a separation, or bubble ("blister") in fundamental spacetime geometry.

Penrose (continued):

Superposed quantum states for which the respective mass distributions differ significantly from one another will have space-time geometries which correspondingly differ. Thus, according to standard quantum theory, the superposed state would have to involve a quantum superposition of these differing space-times. In the absence of a coherent theory of quantum gravity there is no accepted way of handling such a superposition. Indeed the basic principles of Einstein's general relativity begin to come into profound conflict with those of quantum mechanics (cf. Penrose, 1996). Nevertheless, various tentative procedures have been put forward in attempts to describe such a superposition. Of particular relevance to our present proposals are the suggestions of certain authors (i.e., Karolyhazy, 1996; 1974; Karolyhazy et al., 1986; Kibble, 1991; Di'si, 1989; Ghirardi et al, 1990; Pearle and Squires, 1995; Percival, 1995; Penrose, 1993; 1994; 1996) that it is at this point that an objective quantum state reduction (OR) ought to occur, and the rate or timescale of this process can be calculated from basic quantum gravity considerations. These

particular proposals differ in certain detailed respects, and for definiteness we shall follow the specific suggestions made in Penrose (1994; 1996). Accordingly, the quantum superposition of significantly differing space-times is unstable, with a lifetime given by that timescale. Such a superposed state will decay - or "reduce" - into a single universe state, which is one or the other of the space-time geometries involved in that superposition.

Whereas such an OR action is not a generally recognized part of the normal quantum-mechanical procedures, there is no plausible or clear-cut alternative that standard quantum theory has to offer. This OR procedure avoids the need for "multiple universes" (cf. Everett, 1957; Wheeler, 1957, for example). There is no agreement, among quantum gravity experts, about how else to address this problem.

So Penrose OR is a self-organizing process in fundamental spacetime geometry. If that is where qualia and Platonic values are embedded, an OR process occurring in the brain could be the connection.

Figure 9. Quantum coherent superposition represented as a separation of space-time. In the lowest of the three diagrams, a bifurcating space-time is depicted as the union ("glued together version") of the two alternative space-time histories that are depicted at the top of the Figure. The bifurcating space-time diagram illustrates two alternative mass distributions actually in quantum superposition, whereas the top two diagrams illustrate the two individual alternatives which take part in the superposition (adapted from Penrose, 1994 - p. 338).

Figure 10. Superpositions and qubits (qubits are information "bits" which can exist and compute while in superposition, then reduce to a classical output state). A. Protein qubit. A protein such as tubulin can exist in two conformations determined by quantum London forces in hydrophobic pocket (top), or superposition of both conformations (bottom). B. The protein qubit corresponds to two alternative spacetime curvatures (top), and superposition/separation (bubble) of both curvatures (bottom).

In conclusion, Penrose OR may be the connection between the brain and 'fundamental' reality. Consciousness may involve self-organization of spacetime geometry stemming from the Planck scale.

The question of whether such a process could actually occur in a biological environment will be discussed in future lectures.

Browne, M.W. 1997. Physicists confirm power of nothing, measuring force of universal flux. The New York Times, January 21, 1997.

Chalmers, D. (1996) Facing up to the problem of consciousness. In: Toward a Science of Consciousness - The First Tucson Discussions and Debates, S.R. Hameroff, A. Kaszniak and A.C. Scott (eds.), MIT Press, Cambridge, MA.

Chalmers, D. (1996) *Toward a Theory of Consciousness*. Springer-Verlag, Berlin.

Conze, E., (1988) *Buddhist Thought in India*, Louis de La Vallee Poussin (trans.), Abhidharmako'sabhaa.syam: English translation by Leo M. Pruden, 4 vols (Berkeley) pp 85-90.

Diosi, L. (1989) Models for universal reduction of macroscopic quantum fluctuations. *Phys. Rev. A*. 40:1165-1174.

Everett, H., (1957) Relative state formulation of quantum mechanics. In *Quantum Theory and Measurement*, J.A. Wheeler and W.H. Zurek (eds.) Princeton University Press, 1983; originally in *Rev. Mod. Physics*, 29:454-462.

Goswami, A., (1993) *The Self-Aware Universe: How Consciousness Creates the Material World*. Tarcher/Putnam, New York.

Hall, G.L. 1996. Quantum electrodynamic (QED) fluctuations in various models of neuronal microtubules. *Consciousness Research Abstracts -Tucson II (Journal of Consciousness Studies)* Abstract 145.

Ingber D.E. 1993. Cellular tensegrity: Defining new roles of biological design that govern the cytoskeleton. *J Cell Science* 104(3):613-627.

Jibu, M., Yasue, K. 1995. *Quantum brain dynamics: an introduction*. John Benjamins, Amsterdam.

Jibu M, Pribram K.H., Yasue K 1996. From conscious experience to memory storage and retrieval: The role of quantum brain dynamics and boson condensation of evanescent photons. *Int J Modern Physics B* 10 (13&14):1735-1754.

Jibu, M., Hagan, S., Hameroff, S.R., Pribram, K.H., and Yasue, K. (1994) Quantum optical coherence in cytoskeletal microtubules: implications for brain function. *BioSystems* 32:195-209.

Pearle, P. (1989) Combining stochastic dynamical state vector reduction with spontaneous localization. *Phys. Rev. D*. 13:857-868.

Pearle, P., and Squires, E. (1994) Bound-state excitation, nucleon decay experiments and models of wave-function collapse. *Phys. Rev. Letts*. 73(1):1-5.

Penrose, R. (1987) Newton, quantum theory and reality. In *300 Years of Gravity* S.W. Hawking and W. Israel (eds.) Cambridge University Press.

Penrose, R. 1971. in *Quantum Theory and Beyond*. ed E.A. Bastin, Cambridge University Press, Cambridge, U.K.

Penrose, R. (1989) *The Emperor's New Mind*, Oxford Press, Oxford, U.K.

Penrose, R. (1994) *Shadows of the Mind*, Oxford Press, Oxford, U.K.

Penrose, R. (1997) On understanding understanding. *International Studies in the Philosophy of Science* 11(1):7-20, 1997.

Rovelli C, Smolin L 1995b. Spin networks in quantum gravity. *Physical Review D* 52(10):5743-5759.

Seife, C. 1997. Quantum mechanics. The subtle pull of emptiness. *Science* 275:158.

Shimony, A., 1993. *Search for a Naturalistic World View Volume II. Natural Science and Metaphysics*. Cambridge University Press, Cambridge, U.K.

Tart, C.T., (1995) personal communication and information gathered from "Buddha-1 newsnet"

von Rospatt, A., (1995) *The Buddhist Doctrine of Momentariness: A survey of the origins and early phase of this doctrine up to Vasubandhu* (Stuttgart: Franz Steiner Verlag).

Wheeler, J.A. (1990) Information, physics, quantum: The search for links. In (W. Zurek, ed.) Complexity, Entropy, and the Physics of Information. Addison-Wesley.

Lecture 2

Interpretations of quantum mechanics and the nature of reality

Classical physics, as it had developed up to the end of the 19th century, saw that there are two basic kinds of entities in the universe, particles and fields. The particles were thought to follow Newton's laws of motion, while the fields were thought to obey Maxwell's equations for the electromagnetic field. Lord Kelvin said that physics was pretty much finished except that there were two small clouds in the horizon, the negative results of the Michelson-Morley experiment (the search for ether) and the failure of Rayleigh-Jeans law to predict black-body radiation. Lord Kelvin chose his clouds well for the former gave rise to relativity and the latter to the quantum theory.

The ontologically essential lesson of the quantum theory was that the classical idea of particles and fields was wrong. The electromagnetic field turned out to have a particle aspect, and particles like the electron turned out to have a field aspect. The most fundamental ontological feature of the quantum theory then is that each manifestation of matter and energy can have two possible aspects, that of a wave and that of a particle. This is philosophically significant because it is a universal feature. The debate about the interpretation of the quantum theory has to do with how to deal with this observed fact of wave-particle duality.

Quantum theory began when it was discovered that the electromagnetic field has a particle aspect in the sense that it exchanges energy with particles in small discrete amounts or quanta. Thus light (previously thought to be a form of wave motion due to observed diffraction and interference effects) was thought to consist of particles or photons with energy $E = hf$ when it interacts with matter (h is Planck's constant and f is frequency of the wave aspect of the light). This is paradoxical because the energy of a bullet of light is expressed in terms of frequency which is a wave property. So a light bullet is no ordinary bullet. The energy of the particle aspect of light is determined by the frequency of its wave aspect. This is completely against anything that was encountered in classical physics. Particles like electrons were initially thought to be like little bullets which circulate around the nucleus pretty much like the planets circulate around the sun. It was possible to measure typical particle properties like mass and charge for the electron. However, it was discovered that the electron - and all subatomic particles - also have a wave aspect.

This can be clearly seen in the well-known two-slit experiment. (If you are not familiar with this, check out the brilliant web site of the University of Colorado:

<http://210.107.224.101/physics2000/index.html> go to the atomic lab and choose the two-slit experiment.)

Because the quantum theory is often presented in the spirit of the Copenhagen interpretation which emphasizes indeterminacy and probability, people sometimes overlook very simple facts regarding the wave properties of the electron. Note first that the interference pattern which is built up spot by spot is very determinate - so there is "statistical causality" at the quantum level, not just indeterminacy of the behaviour of the individual system. The key question to ask is why we get an interference pattern rather than some other pattern or no definite pattern at all (which one might expect if electrons were truly indeterminate in their behaviour). Classically we would expect just "two piles" of electrons behind the screen, rather than the observed interference fringes or "many

piles" with regions of no electrons between them. It is just a basic experimental fact that we get an interference pattern which can be predicted with the mathematics of wave motion. The reasonable way to think about this is to say that there is a wave aspect associated with each electron and that this wave aspect is causally powerful in that it makes sure that the particle aspect of each individual electron obeys the interference pattern, never going to the "forbidden regions".

However, one must watch out not to think of this wave as just an ordinary wave of classical physics. For one thing for the many-body system the wave lives in a $3N+1$ dimensional "configuration space", where N is the number of systems that we are considering and the additional dimension is time. Also, because of the uncertainty principle it is not possible to observe how an individual electron is moving under the guidance of the wave, and thus the idea of such movement and guidance is thought often to be "metaphysical" - a mere hypothesis without the possibility for experimental verification. The standard view has been to think of the quantum wave as a kind of probability wave.

The participants of this course have no doubt very different backgrounds in their knowledge of quantum mechanics. The world wide web provides a good resource for introductions and visualizations of the various ideas. One good brief description of the standard view of quantum mechanics is the piece "Some Basic Ideas about Quantum Mechanics by Stephen Jenkins at

<http://newton.ex.ac.uk/people/jenkins/mbody/mbody2.html>

and it is recommended that you read this at this point if you are not familiar with the standard view. For a more technical account of the Copenhagen interpretation of quantum mechanics, check John G. Cramer at

http://mist.npl.washington.edu/npl/int_rep/tiqm/TI_20.html

Non-locality vs. Non-reality

The EPR Paradox and Hidden Variables

In one version of the famous thought experiment due to Einstein, Podolsky and Rosen (1935), a neutral pion (π^0) decays into an electron (e^-) and a positron (e^+):

$$\pi^0 \rightarrow e^- + e^+$$

Because the neutral pion is spinless, quantum mechanics prescribes that the two decay products must have opposite spin. Until a measurement is made on one member of the duo, say the electron, there is equal probability that the measurement will yield an "up" or a "down" result. However measuring "up" spin for the electron then implies that the positron *must* have "down" spin; measuring "down" spin for the electron implies that the positron *must* have "up" spin. And this must hold regardless of how far apart the electron and positron were when the measurement was made. Experiments of this kind have been performed in actuality (Aspect, Grangier and Roger, 1982; Aspect, Dalibard and Roger, 1982) and the results have proven to be completely consistent with quantum mechanics.

Figure 1 [EPR] One possible realization of the Einstein-Podolsky-Rosen *Gedankenexperiment*. Measuring "up" spin for the electron immediately implies that the

positron *must* have "down" spin, and *vice versa*, regardless of how far apart they might be when measured.

Einstein, Podolsky and Rosen (EPR) assumed in their analysis of the thought experiment that no non-local influence could instantaneously inform the positron that a measurement had been made on the electron (after all, it was Einstein who introduced the Theory of Special Relativity that formalized the notion of a locality constraint), so they concluded that the spin of the positron must be a "real" property (a position known as *quantum realism*), determined by variables that had not been or perhaps could not be measured - so-called *hidden variables*. Their contention was that quantum mechanics was "incomplete"; that the probabilistic distribution of spin measurements determined by experiments was a result of our ignorance of these hidden variables. In their view, if we knew the values of the hidden variables, the results of measurements would be given deterministically, just as in classical mechanics.

Von Neumann (1932) was the first to put forward a theorem to the effect that hidden variable theories consistent with the predictions of quantum mechanics were impossible. Working from the *density matrix* formalism that is widely believed to stand or fall with quantum mechanics itself, von Neumann established that no quantum mechanical *ensemble* is *dispersionless*. (A dispersionless ensemble is one for which the square of the average value for any observable is equal to the average value of the observable squared.) Although this theorem could at best prove only that non-dispersive hidden variable theories were impossible, von Neumann believed that such theories were the only kind that had to be eliminated from consideration in order to establish his thesis. Only this kind of hidden variable theory, the non-dispersive kind, could eliminate the statistical element from quantum mechanics and reduce it to the kind of classical theory that would allow one to predict the results of individual acts of observation. But his theorem overlooked the possibility that if the hidden variables were themselves of a statistical nature (dispersive) then, even though the statistical element would not be eliminated from quantum mechanics, it could nevertheless be made compatible with a local, causal description of how one could understand the spin of the electron and positron in the EPR thought experiment. It was not until almost thirty years later that Bell proposed a similar theorem (Bell, 1964, 1966) that made it much more clear what assumptions had to be made about hidden variable theories in order to eliminate them from consideration.

Bell's Theorem

Bell's theorem demonstrated that quantum mechanics was in fact not compatible with hidden variables, at least not if you wanted the hidden variables to be real properties determined locally; that is, if you wanted to interpret hidden variables as having some determinate value regardless of whether or not there was a 'measurement situation' and if you wanted that determinate value to depend only on the 'particle' being measured. To accommodate the predictions of quantum mechanics, as borne out by experiment, either the reality or the locality assumption must be relaxed.

To demonstrate this, Bell starts from the same EPR thought experiment with the modification that the spin, of both the electron and the positron, can be measured along *any* specified direction. Bell's theorem will then be a statement about the average value that one obtains when the two spin measurements, one on the electron and one on the positron, are multiplied.

If the hidden variables can be given *local* instantiation then they can be assigned a probability density and the average of the product of spin measurements can be determined in a prescribed way.

If it is true, as asserted by hidden variable theories, that the outcomes of all possible experiments that might be performed on the electron/positron system are *simultaneously* well-defined (as would be the case if these outcomes were decided by hidden variables taking *real* values that were simply unknown to us), then it should be possible to write consistent mathematical statements *simultaneously* invoking both the *actual* measurement of the positron spin and *counterfactual* measurements, ones that *might* have been made but were not. Thus the mathematical statement that Bell derives incorporates the EPR reality assumption that the result of a measurement is determined whether or not the measurement is made.

Bell's conclusion is that for any theory involving local hidden variables a certain inequality must always hold, an inequality that involves the average product of electron and positron spin measurements made under various alignments of the measurement apparatus, at least some of which must be counterfactual. The actual inequality is rather cumbersome regardless which of several different formulations one chooses. One of the possible versions (Clauser *et al.*, 1969) states that the positive difference between the average products calculated when the measurement devices (for electron and positron respectively) are aligned along directions **a** and **b**, and when they are aligned along directions **a** and **c**, must be less than or equal to two *plus or minus* the sum of the average product obtained when the devices are aligned along directions **d** and **c**, and when they are aligned along **d** and **b** (directions **a**, **b**, **c** and **d** being arbitrary). A mouthful, to be sure.

Figure 2 [Bell] In the formulation due to Clauser *et al.* (1969), Bell's theorem can be given this graphical representation, where each depiction of an EPR-type experiment represents the average product of electron and positron spins calculated for the alignment of detectors shown. For particular choices of the alignments **a**, **b**, **c** and **d**, Bell's inequality is violated by the predictions of quantum mechanics.

In particular cases this statement, derived under the assumption that the hidden variables of EPR must be both real and local, can be shown to be in contradiction with the predictions of quantum mechanics. Either quantum mechanics is wrong or there is no room for local hidden variables.

The experiments of Aspect *et al.* (Aspect, Grangier and Roger, 1982; Aspect, Dalibard and Roger, 1982) have shown the predictions of quantum mechanics to be correct in cases where a violation of Bell's inequality is incurred. The experimental design made it possible to ascertain that this result holds even in situations where the particles being measured (these were actually two correlated photons rather than an electron and a positron, but the principle is the same) were separated such that no possible signal could have informed the second particle of the measurement result on the first, prior to the second being itself measured.

Since the premise of Bell's theorem, that it is possible to write a theory that gives a complete description of the state by introducing hidden variables that are instantiated locally, leads to a false conclusion, we have a *reductio ad absurdum* indicating that the premise must also be false.

Whether the premise is false because it assumes the *locality* of the variables or because it assumes the *reality* of the variables is currently a topic of lively debate. Stapp (1997) has, for instance, introduced an argument that, if correct, would indicate that it is in fact the locality assumption that is at fault, for he derives a contradiction similar to that of Bell, apparently without assuming the reality of any hidden variables. This is a particularly strong claim, one that would herald significant changes in the way physicists think about natural phenomena. Equally though it is one to which there has been great resistance in the physics community since, whenever non-local influences obtain in an experimental situation, the sphere of causal influence can no longer be restricted. (This should not, of course, be taken as any argument against Stapp's contention, since what physicists would like to be true presumably has no influence on what actually is.) For this reason considerable controversy surrounds Stapp's claim.

At issue is the use that the argument makes of counterfactuals (Mermin, 1997; Unruh, 1997; Stapp, 1998). Counterfactual reasoning is ubiquitous in classical science and plays a recognized role in theorizing about the quantum domain as well, however the principles of quantum mechanics must preclude certain uses that would be valid in classical reasoning and it is not a matter of general agreement where the line should be drawn. The concern is that, if counterfactuals have been employed in a quantum mechanical situation to which they would be inappropriate, then a reality condition has been smuggled into the argument implicitly. If this were the case then Stapp's argument would reach the same conclusion as Bell's theorem. One avenue of exploration that may be interesting in this regard is the proposal made by Griffiths (1996, 1998) of a set of conditions - framed in terms of *consistent histories* - that would ensure that orthodox logic could be applied consistently in quantum contexts. In a recent addition to the debate, Stapp (1999) himself has invoked this framework in defence of his argument.

Non-collapse Interpretations

Bohm's interpretation

Bohm's interpretation will be presented in more detail in lecture 5, so here is just a brief summary. Bohm had written a textbook "Quantum Theory" (1951) in which he attempted as far as possible to give a physical formulation for the quantum theory while staying within the then standard "Copenhagen" view. Many physicists like Pauli and Einstein thought that he had done a great job, yet he felt he didn't really understand the theory. One puzzling question was that of ontology. Quantum theory didn't provide a clear ontology, yet by making the so called WKB approximation one could derive classical mechanics out of quantum mechanics, thus also obtaining the straightforward classical ontology. Considering this Bohm saw that if one didn't make the approximation one could obtain a "quantum ontology", where there was an extra potential added to the classical ontology.

Physically this meant that one could imagine the electron as a particle always accompanied by a new kind of wave that gave rise the new "quantum potential" acting on the particle. He went on to show in his well known 1952 Physical Review papers that you can explain basic features of non-relativistic quantum theory by assuming that the

electron is such an entity with two aspects: a particle aspect which explains why we observe a particle-like manifestation every time we observe the electron, and a wave aspect which acts on the particle in a subtle way, thus explaining why the particle aspect obeys the mathematics of wave motion and why electrons collectively produce interference patterns, without the need to assume the collapse of the wave function. The theory was a non-local "hidden variable" theory and was thus a counter-example to von Neumann's proof that hidden variables are impossible. It met initially with great resistance but is today, in various developed forms, considered as one of the serious alternative interpretations of the quantum theory.

Everett's interpretation

Hugh Everett III's interpretation of quantum mechanics (Everett, 1957, 1973), or 'relative state' theory, despite being the interpretation most often celebrated in fiction, has been largely misinterpreted in popular accounts, at least with respect to Everett's original proposal and this has in part been due to the fact that later adherents imposed their own reading of the theory. While some of these are worthy of consideration in their own right, much of the criticism directed against Everett's interpretation has in fact objected to aspects that did not appear in the original version.

Like Bohm's interpretation, Everett's involves no 'collapse of the wave function'. But unlike Bohm, Everett accepts the 'reality' of quantum superposition. The heart of Everett's proposal is in fact that superposition obtains in every case and at every level of description, whether the object of description is the spin state of an electron, the state of a pointer in a measuring instrument or even the state of the experimenter's mind looking at a measurement device.

Everett proceeds from the observation that collapse is not really a necessary element of quantum mechanics if there exists a means to establish rigorous correlations between superpositions at various levels.

Imagine for instance that an electron, existing in a superposition of 'up' and 'down' spin states, is 'measured'. For Everett, this means that the superposition of spin states of the electron leads to a similar superposition in the measuring device, say a superposition of the word 'up' being printed and of the word 'down' being printed. The measurement situation provides an association between the 'up' state appearing in the superposition of electron states and the 'up' state appearing in the superposition of measurement device states. This can be taken even further so that when the experimenter looks at the printed output, her mind also enters a superposition of states involving a state in which she reads the word 'up' and a state in which she reads the word 'down'. Again there is an association between elements of the superposition at the level of the printout and elements of the superposition at the level of the experimenter's mind. From the perspective of any one of these superposed minds however, there appears to be no superposition at all. That is because the conscious state of the experimenter has become entangled with the state of the printout in such a way that the conscious state in which the word 'up' is read is always accompanied by the state of the measurement device in which the word 'up' is printed. Because the mind is not *outside* the superposition, but rather itself superposed, the experimenter exists in one mental state or the other and so sees only the word 'up' *or* the word 'down' rather than a superposition of both.

This has been taken to imply that, since there are now two observers, the universe itself has bifurcated, each of the resulting branch universes accommodating only one of these

observers. The thought behind this is that there is only one 'me' in this universe and any other versions of 'me', that might perhaps remember different results in a quantum mechanics experiment, should go find their own universe. On this reading, every time a 'measurement' is made, the universe must similarly 'split' (hence the designation 'many-worlds interpretation'). It is this ontologically prodigious contention that has been repeatedly criticized as extravagant excess. It gives no account of the process whereby a 'split' might occur and one cannot therefore raise the relevant concerns. Is it appropriate to raise an objection on the basis of energy conservation? Or perhaps with respect to the 'no-cloning theorem' (according to which the information completely specifying a quantum system cannot be copied from that system without destroying the original)? Or simply as regards parsimony? Without further information, the interpretation would appear too under-specified to decide. Even if splitting turns out to be unobjectionable in itself, there nevertheless remains the problem of deciding what exactly should constitute the 'measurement' that provokes the split. The measurement problem that initially prompted the quest for an interpretation of quantum mechanics remains thus an unresolved paradox. All of this however pertains to a reading that cannot be traced to Everett's work and may be largely credited to a sympathetic misconstrual by DeWitt (1970, 1971).

Albert and Loewer (1989; also Albert, 1992), Lockwood (1989) and Chalmers (1996) have urged a reading that eschews the objective splitting of the universe. On this version a single universe evolves a vast superposition: electrons, measuring devices and human minds all added to the mix. Superposition of observers is treated no differently from any other case of superposition and prompts no bifurcation of worlds but the world takes on a classical appearance *relative* to any one of the superposed observers (hence 'relative-state' theory).

One of the abiding difficulties with even this more reasonable version is the problem of *preferred basis*. The relative-state interpretation assumes a basis in which, relative to the mind of an observer, the mutually exclusive outcomes of quantum experiments always show up as discrete classical alternatives. This is not a unique possibility, for the mind of the observer might just as well be correlated with one of two mutually exclusive *superpositions*, in which event the world would not retain its classical appearance. But no justification can be given for denying the possibility without the introduction of new principles.

Neither does the interpretation give us any understanding of the amazing ability of the quantum mechanical calculus to predict the *probabilities* of various outcomes, probabilities that are borne out in statistical trials. In collapse interpretations of quantum mechanics, the squared modulus of each coefficient in the wave function determines the relative probability of realizing the corresponding term upon collapse. Since the wave function in relative-state theory never collapses, it is not clear how one might recover this principle, obviating the need for further assumptions.

Everett himself addressed the problem by introducing a 'measure' over the space of observers. This measure weights each branch of the wave function with the standard quantum mechanical probability, but the measure is explicitly introduced as an abstract and uninterpreted quantity, and specifically not as a probability measure. Everett goes on to show that the set of observers who would *not* find a statistical distribution of results consistent with quantum theoretical predictions, has measure zero, the implication being

that such observers will never occur. As pointed out by d'Espagnat (1976), the implication follows however, only if the measure *is* interpreted in terms of probability. The measure then appears as either an *ad hoc* adjustment to square the theory with the predictions of quantum mechanics or, minimally, as a further assumption. Since no objective realization of the measure is supplied, there is no way to argue that the implication follows from the significance of the measure.

In an attempt to give objective meaning to the measure and to cement the implication, Albert and Loewer, in their 'many-minds' version, interpret the weighting as occurring in a space of *actual* minds. They begin in a manner reminiscent of DeWitt's 'many-worlds' view, but assume that it is not worlds that split but minds, and that the branching minds are all unaware of each other. This shift allows them room to *define* minds such that superposition fails as applied to them. Minds are ideal candidates as they appear never to be in superposition. It is thus the splitting of minds that, in an over-all random process, is weighted by the quantum mechanical probabilities, so that the probabilities are realized as relative proportions of an ensemble of minds that, according to the circumstance, may be vast or even infinite in number. The 'many-minds' picture thus avoids the problem of preferred basis - minds sit outside the principle of superposition, and themselves define the favored basis in which the split occurs - but as in the DeWitt interpretation, we have again a problem of extravagance. If it is already a problem to explain how *one* conscious state might arise from *one* brain state, then the problem is clearly compounded by imagining that *one* brain state gives rise to a vast, potentially *infinite*, number of conscious states.

Collapse and 'Collapse-like' Interpretations Decoherence

The theory of decoherence was originated by Murray Gell-Mann and James Hartle (1989) to contend with problems surrounding the interpretation of quantum mechanics in a specifically cosmological context. While cosmology may not be an evident concern in consciousness studies, the theory of decoherence has attained considerable prominence with respect to certain of the proposed models. In reviewing this interpretation, it will be important to bear in mind its origins in cosmology and how that has shaped it in a specific image.

Cosmologically, it is difficult, if not impossible, to maintain the line between 'observer' and 'observed' postulated in those interpretations that make a sharp "cut" between quantum and classical. That is, we do not expect, given the extreme nature of the ambient conditions immediately following the Big Bang, that the early universe was inhabited by nascent consciousnesses. Particularly if we are to take seriously some form of cognitive coherence - that consciousness obtains only where it is correlated with a physical entity capable of sensing, cognizing and reacting to its environment - the postulation of primordial observers seems clearly problematic. Observers, at least biological observers, would seem to require perhaps billions of years of interceding evolution to facilitate their appearance on stage. Are we to imagine then that until such time as biological entities evolved to the level of consciousness that the entire universe remained in perpetual superposition?

Wheeler (1957), at one time, put forward exactly this hypothesis: that the appearance of conscious observers in the superposition of all possibilities collapsed the wave function of the universe retroactively, thus *creating* in a sense the entire classical history of all that

is. While perhaps intriguing from the perspective of recursion and certainly grand in scope, the hypothesis repudiates key features of the formulation of quantum mechanics. The 'observer' in Wheeler's proposal is not outside the phenomenon 'observed' - conscious observers appear *within* the superposition - so the divide between classical and quantum, a divide initiated by Bohr and carried through von Neumann, Heisenberg, Dirac and Wigner, has been abandoned. This means that a quantum system must be able to initiate its own collapse, but if this can be facilitated then there may be no reason to single out consciousness as a factor. We may find a route more amenable to expression in the context of standard physical theory.

Decoherence is an attempt to do just that. 'Measurement' is regarded as a specific case of a process that is going on all the time, effecting the reduction of quantum to classical without the necessary involvement of conscious entities.

Gell-Mann and Hartle (GH) build a theory of *decohering histories*. They attempt to show that, in what they call a *quasi-classical* domain, a quantum superposition of histories can be made to decohere to a single, approximately classical history. Each possible history is assigned a probability of being actualized quasi-classically but the theory must nevertheless preserve the quantum superposition of histories wherever coherence is a critical factor in the explanation of empirical results (as in the two-slit experiment, for instance).

Histories are defined in the theory in terms of *projection operators*. These operators, acting on state vectors, project out particular properties at a given time. At any particular time, a set of projection operators that covers the whole spectrum of possible values for some set of fundamental field variables is called *complete*.

We might then go on to construct *alternative histories* as time sequences of these projection operators, by choosing from amongst the possible complete sets at each moment. This sense of the word 'history' is analogous to the sense used by Feynman in his *sum over histories* approach to quantum mechanics. There, a 'history' specifies one of the many possible routes that an electron (for example) might take between an initial and final measurement. In the theory of decohering histories, we cannot allow 'measurements' to enter as elementary entities; 'measurements' are what decoherence theory seeks to eliminate. For this reason a 'history' in the sense of decoherence theory must necessarily involve the entire evolution of the universe, a fact appreciated by Gell-Mann and Hartle themselves. They will thus average their alternative histories over the initial *density matrix of the universe*, which describes the quantum state of the very early universe. Any later state from which they might commence a set of 'histories' would be specified in virtue of a 'measurement', a process that must be expunged from the theory. (Moreover, whatever density matrix is to be used as a starting point must satisfy a highly restrictive set of conditions on the orthogonality of alternative histories - conditions that, while perhaps acceptable in the early universe, are acknowledged by GH to be too limiting at later times.)

The approach outlined leaves out one crucial detail of the Gell-Mann and Hartle proposal, and with that omission, runs aground immediately. As we have so far defined our alternative histories - in terms of *complete* sets of projection operators - it will not prove possible to consistently assign probabilities to histories in a way that will be compatible with the known results of quantum mechanics where its laws hold sway. To escape this impasse, GH effect a *coarse-graining* procedure on the projection operators.

Rather than being defined for specific *values* of the fundamental field variables (a *fine-grained* approach), the projection operators are defined over a certain *range* of values. Rather than phrasing their alternative histories in terms of *complete* sets of operators, they work with *exhaustive* sets - those for which the total set of *ranges* covers all possible values. Their histories are then composed by selecting from these exhaustive sets of operators at each moment in the sequence. With coarse-graining, GH claim that it is possible to consistently assign probabilities to alternative histories.

Information about which histories will decohere is contained in the *decoherence functional* (a functional is, loosely speaking, a function of functions; more precisely, it maps functions to numbers), which will involve, in addition to the two alternative histories, the initial density matrix of the universe. When the decoherence functional returns a value of zero for two alternative histories, these are considered to have decohered (that is, they will correspond to alternative possibilities in the quasi-classical world).

GH make the further claim that when the value is *small* these histories might still be regarded as having decohered. In this case, the rules for assigning probabilities to alternative histories are only approximately respected, and the size of the non-zero value for the decoherence functional measures the failure of that approximation. Adherence to these rules is imperative if one is to obtain in the appropriate circumstances the standard results of quantum mechanics. Coarse-graining has a tendency to make the decoherence functional of distinct histories small and so the coarser the grain, the less tension is observed between decoherence theory and the predictions of quantum mechanics. On the other hand, the grain cannot be made so coarse that the quasi-classical world becomes ambiguous as no such ambiguity is observed classically.

Ultimately the theory will require an objective scale parameter that sets the level of coarse-graining and thereby determines at what point the quantum regime breaks down. That parameter will have to be adjustable to allow, for instance, for the fact that systems at high temperature can accommodate a much finer grain than systems at low temperature.

How do GH explain the incommensurability of particular quantum variables, say position and momentum? In their view the decision of an experimenter to set up a measurement correlation with one or the other of these variables is an intimate part of the history to which it belongs, a history that must fold in all of the factors influencing the experimenter's memory or awareness, in fact the entire chain of events leading from the initial state of the universe to the moment of the experiment. In a measurement situation the alternatives that correspond to measuring either position or momentum are represented by decohering histories and are thus mutually exclusive.

This immediately entails that the experimenter is not freely choosing, and may further imply what has been called the "denial of the counterfactual" (D'Espagnat, 1976) - that, given the history of the universe such as it has been, there were not in fact different counterfactual possibilities for how it *might* have proceeded. The outcome was as it had to be, as set out from the time the universe was born. If this charge is well-founded then the validity of decoherence theory would argue against the soundness of any results that depend on a counterfactual element, such as does Bell's theorem. If counterfactuals do not exist, they could not be admitted as legitimate theoretical devices.

Equally serious is the apparent threat to separability, one of the foundations upon which science has historically been based. It has for hundreds of years been the desideratum of the scientific community that experiments should be amenable to analysis in relative isolation from the rest of the universe and the course of its evolution (that these should be completely subsumed in the boundary and initial conditions to which the situation is subject).

In part to address these concerns, to reduce the scope of inquiry, recent work in decoherence theory has focussed on dynamic decoherence (for early work, see for example Caldeira and Leggett, 1983; Zurek, 1984)- decoherence that results as a dynamic consequence of interactions with an environment. Foundational issues aside, this approach has made considerable progress in characterizing the evolution of simple systems from the quantum to quasi-classical regime, without involving the early universe. These studies envisage such a system, for instance a driven anharmonic oscillator, as being coupled to an 'environment' consisting of a large number of oscillators initially in thermal equilibrium. What the research has been able to show is that the alternative histories associated with different trajectories of such systems can rapidly decohere in the presence of a significant thermal background. This work suggests that thermal noise might effect decoherence on 'real world' systems in short times.

Stapp

Stapp (1986) is strongly influenced in his formulation of an interpretation of quantum mechanics by the concept of 'process' introduced in Whiteheadian philosophy. In 'process' the primitive notion of 'states', instantaneous elements of actuality having themselves no duration, is replaced by what Whitehead calls 'occasions' and Stapp refers to as 'events'. Each event makes reference to a bounded space-time region rather than an instantaneous time slice and a process is a well-ordered sequence of such events.

To each event there corresponds a restriction on the possibilities that might be actualized within the relevant bounded space-time region. Stapp frames this restriction, in a manner reminiscent of the approach adopted by GH above, through the use of projection operators (however, note that Stapp's proposal predates GH). These operators restrict the set of all conceivable fields to those that might actually be realized in the corresponding space-time region. Unlike standard quantum theory, where projection operators are indicative of the possible results of a 'measurement situation', projections are here taken to be constitutive of an event and do not necessarily imply measurement. A sequence of successive projections, averaged over the initial density matrix of the universe, can be used to determine the density matrix for the set of realizable possibilities, and from this probabilities can be calculated.

To the extent that the superficial similarities to GH's decoherence approach carry through, Stapp's proposal raises similar questions. Like GH, Stapp must ultimately provide a criterion on the basis of which the degree of coarse-graining appropriate to the projection operators might be judged. Presumably this criterion would be implicit in a complete definition of an 'event', a concept that greatly distinguishes the character of Stapp's theory from that of decoherence. In developing this concept, Stapp has made the suggestion that an event should ultimately be defined in terms of consciousness. Of this we shall hear more in later lectures.

Ghirardi-Rimini-Weber

The proposal of Giancarlo Ghirardi, Alberto Rimini and Tullio Weber (GRW, 1986) is one of several which attempt to realize 'collapse', or something physically similar, as an objective phenomenon (Károlyházy, 1966; Károlyházy, Frenkel and Lukács, 1986; Diósi, 1989). In this scheme, the evolution of the wave function is assumed to proceed according to the usual (linear) Schrödinger equation but with the addition of a new, non-linear term that subjects the wave function to occasional, random localizations in configuration space. Under ordinary Schrödinger evolution, a wave function, initially well-localized, has a tendency to spread out. GRW's addition is a universal process that acts to increase its localization, and thus its classical appearance. This might occur, following their suggestion, by multiplying the wave function by another, highly-peaked function, like a Gaussian (bell curve).

Figure 3 [GRW]: In the GRW scheme the usual tendency for a wave function to spread out is counteracted by a random, spontaneous localization procedure that multiplies the wave function by a Gaussian. The frequency with which this process takes place is proportional to the number of entangled particles and so occurs much more often for objects in the macroscopic domain.

This accomplishes several things. First of all, it preserves the idea of determinate dynamics developing according to a Schrödinger-like equation. And because particle wave functions are universally and randomly susceptible to the localization process, GRW are able to dispense with the problematic notion of 'measurement' and all its attendant difficulties. In fact, their proposal contains no actual collapse so that even a macroscopic wave function always remains in superposition, albeit one that, following localization, has its amplitude highly concentrated in one area.

GRW must however show that this idea can be made consistent with the known results of quantum mechanics for microscopic systems but nevertheless, at macroscopic scales, quickly lead to a situation qualitatively indistinguishable from the classical case.

Correspondence with the microscopic domain is achieved by setting the characteristic frequency with which the localization procedure occurs - how often, on average, one expects the multiplication of the wave function by a Gaussian to take place. In the GRW scheme, this is determined by a new universal constant to which they give an approximate value of 10^{-16} per second, or about once every billion years.

For macroscopic objects, localization will have to occur *much* more rapidly. At this level the wave functions of individual particles are expected to be *entangled* (that is, they are not separable). GRW argue from this fact that the frequency of localization in large objects will be proportional to the number of particles. Since the number of particles in even very small macroscopic objects is so large, this leads to rapid localization and the figures obtained appear to be in accord with expectations: a water droplet one millimeter in radius for instance should localize in less than a tenth of a millisecond. We then see how the process of *amplification* involved in a measurement situation, by entangling the state of the microscopic quantum system being measured with the macroscopic state of a measuring device, promotes rapid localization.

While macroscopic dynamics are neatly derived from the microscopic, the accordance with classical dynamics is purchased at a price. The GRW scheme introduces two new fundamental constants: the first being the characteristic frequency and the second, the 'spread' of the Gaussian, which GRW set at about 10⁻⁵ cm. The values of these constants are determined merely to accord with the data and have no other justification. They thus appear rather arbitrary. Diósi (1989) has tried to remedy this by fixing the values of the constants in terms of a more fundamental quantity, Newton's gravitational constant, and Penrose (1989) adopts a relation to gravity in his proposal as well.

The procedure envisioned by GRW also incurs a violation of energy conservation that, though small, occurs each time a wave function is multiplied by a Gaussian. Further, a GRW analysis of EPR-type situations, for example, will generically involve non-local energy transfer (the detection of either electron or positron will instantly localize the entire wavefunction despite a potentially enormous spatial spread). Of course, no superluminal transmission of signals, and therefore no violation of special relativity, is thereby implied.

Penrose

Like GRW, Roger Penrose takes 'collapse' to be a close approximation to an objective physical phenomenon, and also like GRW, imagines a process that takes place much more rapidly in the macroscopic domain. But unlike GRW, Penrose proposes a theoretical justification for his criterion in the vehicle of 'quantum gravity' (Penrose, 1987, 1989).

Penrose cites the non-local transfer of energy involved in GRW's scheme as evidence pointing in the direction of quantum gravity. Gravity is invested in the very fabric of space-time, its curvature, in a way that precludes the possibility of defining it locally. While this is not currently well-understood, Penrose contends that non-local features of quantum theory and of gravity implicate each in the explanation of the other. Penrose further notes that only gravity, of all "physical qualities," influences the causal relations between events and that even the concepts of 'energy' and 'time' cannot be defined in a general gravitational context so as to be consistent with the requirements of basic quantum description.

While these assertions are suggestive, Penrose's most powerful intuition comes from a thought experiment (Penrose, 1989) that he refers to as Hawking's Box (Penrose's use of the '*gedanken*' equipment' is unrelated to Hawking's reasons for introducing the situation). It is essentially a universe in a box. The box may not actually be the same size as the universe (there will however be a certain minimum mass that is somewhat larger than the mass of our sun, and the volume must be neither too small nor too large) but everything in the box *is* completely contained, so that nothing, not even the influence of gravity, extends beyond the walls. The box is thus a world unto itself.

As the dynamics of this 'toy world' play themselves out, we will be interested in characterizing the state of the Hawking box in terms of *phase space*. Phase space is a *very* high-dimensional space in which the state of the entire system is given as a single point encoding the positions and momenta of every particle in that system. The figure shows an extremely schematic representation of this, reduced to only two dimensions.

Figure 4 [Hawking Box 1] The phase space of the Hawking Box can be divided between states in which the box contains one or more black holes and states in which the box contains none. The red region encompasses all the states, involving no black holes, that are in thermal equilibrium. The blue region indicates states in which a black hole (or holes) is immersed in a thermal background. Arrows indicate how states in each part of the phase space evolve toward other states.

Over long periods of time (that is, vastly longer than the current age of the universe), we expect one of two situations to dominate. Either things will settle down into a state of thermal equilibrium involving no black holes (all of the states encompassed in the red region) or a state in which one, or perhaps many, black holes are immersed in a thermal background (all of the states encompassed in the blue region). States in between (white regions) are non-equilibrium states (like the current state of the universe, for instance). For each state of the system (that is, each point in phase space), we can ask what state it will move to in the next moment, and because there is always a determinate answer for this in classical mechanics (we're not at quantum theory yet), we could draw an arrow for each point in phase space pointing in the direction of the state to which it will evolve. If we do this for every point we get something like a 'flow diagram', showing how the system would move from one state to the next. If we think of this as a 'fluid' flowing through phase space, then one of the most robust theorems of classical mechanics (*Liouville's theorem*) tells us that this fluid should not accumulate in any part of phase space, that it acts like an incompressible fluid.

This would seem, however, to encounter problems specifically related to black holes. While it seems clear that states with no black holes might occasionally evolve to states involving black holes (as massive stars play out their nuclear fuel, for instance), it does not seem obvious that it should be possible to proceed from a state *with* black holes to one with *none*. But if it were not possible then the 'phase space fluid' would eventually accumulate amongst states involving black holes. That it is in fact possible for black holes to 'evaporate' was the crux of Hawking's most famous contribution to cosmology (Hawking, 1975) and a phase space catastrophe is thus avoided.

But black holes not only swallow up matter and energy, but also information - information that would otherwise specify two different states in the phase space of the box. This implies that, in the part of phase space containing black holes, we should expect to find several distinct flow lines converging into a single flow line as the black hole destroys the information that distinguishes them. In terms of the 'fluid' analogy, this amounts to *destroying* fluid. The incompressible fluid flow is then in trouble unless there is some counter-balancing measure that *creates* fluid, or in other words, accounts for diverging flow lines. But this would mean that a determinate state can evolve in different ways, a circumstance unheard of in classical mechanics but one that is certainly familiar in quantum theory. The measurement process offers several distinct possibilities for the evolution of a system. Which is realized is determined only probabilistically. Penrose thus sees state vector reduction as the flip-side of the destruction of information in black

holes and concludes that the correct interpretation of quantum mechanics must involve gravity.

Figure 5 [Hawking Box 2] If information is actually destroyed in black holes then some of the flow lines in phase space must converge as the information distinguishing different states is lost. If Penrose's conjecture is correct, this should imply that it is also possible for flow lines to diverge - that is, for a single state to have several possible evolutions. Penrose identifies this with the process of state vector reduction and thereby draws his link between quantum theory and gravity.

Penrose admits that his thought experiment does not constitute a full-fledged proof. For one thing, phase space is a classical concept, so if we are to be consistent, the introduction of quantum theory into the thought experiment dictates that the discussion should have been framed from the beginning in terms of Hilbert space, the quantum analogue of phase space. In Hilbert space, the analogue of Liouville's theorem derives from considering the unitary evolution of wave functions. But this unitary evolution presumably only governs the process occurring *between* measurements; state reduction is not a unitary process. The case might therefore not be so strongly stated if it relied on the quantum analogue of Liouville's theorem instead.

There are also issues of principle. The idealizations involved in the thought experiment, the box's utter isolation from everything outside for instance, might contravene some underlying physical law, invalidating Penrose's conclusion. While Penrose is inclined to take such objections seriously, he points to the consistency between results obtained from other Hawking Box experiments (thought experiments unrelated to his own) and those obtained through more 'conventional' channels. Ultimately both sides of the argument reduce to contention and belief but Penrose is unshaken in the conviction that the thought experiment is at least suggestive of a connection between state vector reduction and gravity.

None of this constitutes a concrete theory, but Penrose goes on to make a specific proposal connecting quantum theory with gravity (Penrose, 1993, 1994). He imagines that all quantum superpositions (those that will reduce at least) involve states in which the masses of particles are located in different places. Think of an electron entering a Stern-Gerlach apparatus in which its spin will be measured. If collapse is not something that occurs instantaneously then, at least for a brief interval, the electron wave function exists in a superposition of two states: one in which the electron begins to angle its path in one direction and another in which the electron angles in the opposite direction.

Because it is mass that determines the curvature of space-time, we can interpret this superposition as a superposition of two slightly different space-time curvatures. Penrose postulates that this superposition of space-times creates a tension that builds to a point at which nature finds it intolerable and 'reduces' to a single space-time curvature, selected from the superposed possibilities.

To estimate the time that would be required to bring about this reduction, Penrose calculates how much gravitational energy would be required to 'move' a given mass from

its position in one state of the superposition to its position in the other. The greatest part of this energy is contributed in the early stages, when the separation of the masses is roughly of the same order of magnitude as what we might think of as the 'size' of the particle. Extending the separation, even to infinity, makes little difference to the result. The reciprocal of this energy, measured in appropriate units, gives him an estimate of the reduction time and the numbers are reassuringly of reasonable magnitude. For a proton, for instance, the time is about ten million years (taking its 'size' to be the scale of strong interactions, about 10-15 m) whereas a water droplet a mere micrometer in radius will reduce in only a twentieth of a second. Penrose uses the good accord between these numbers and intuitions regarding the time of collapse to answer critics who assert that the effects of quantum gravity should not be felt on a scale as large as the one relevant to state vector reduction. These critics claim that such effects should occur on length scales given by the Planck length (10-35 m, twenty *orders of magnitude* smaller than a hydrogen nucleus). Penrose counters that, because gravity is so weak compared to the other forces of nature (forty orders of magnitude weaker than electromagnetism for instance), the relevant length scale is actually much larger and becomes an appropriate measure in terms of which to delineate the boundary between quantum and classical. The proposal does not elaborate a mechanism by which reduction actually takes place, nor does the concordance with intuition constitute incontrovertible evidence, but the postulated link to quantum gravity turns out to be one that is amenable to empirical investigation. In a subtle experiment proposed by Anton Zeilinger's lab in Innsbruck, it should be possible to test the hypothesis, although realizing the rather delicate experimental conditions may take several years.

Figure 6 [Penrose]: In an experiment to test Penrose's interpretation, a single photon beam (Incoming Beam) impinges on a half-silvered mirror (Beam-Splitter) and enters a coherent superposition of reflected and transmitted components. The reflected part of the beam enters a cavity resonator (Cavity Resonator I) where it is held for perhaps a tenth of a second while the transmitted part is reflected from a rigid crystal, containing some 10¹⁵ nuclei, into a second cavity resonator (Cavity Resonator II). This places the total mass of the crystal in a superposition of states: one in which it is unmoved and another in which it is very slightly displaced by the momentum of the photon. If the displacement of mass in the superposition is sufficient to satisfy Penrose's criterion, the total wave function encompassing both the reflected and transmitted parts of the photon, and the superposed states of the crystal, must collapse. In such an event, when the beams are routed back over their original paths (a lossless spring must restore the crystal to its original position, without performing a measurement) and recombined at the Beam Splitter, they will no longer be coherent and it will be possible to detect, with a 50% probability, a signal in the path indicated as the Outgoing Beam. If, on the other hand, the superposition remains coherent, the recombined photon *must* pass along the same path on which it entered (Incoming Beam) and no signal will be detected along the path marked Outgoing Beam.

Penrose further believes that the process of objective reduction is not simply random, as in standard interpretations, but rather non-computable (that is, there is *no* algorithm that could even conceivably be used to predict the state that follows objective reduction). For this assertion, Penrose cites tentative evidence in favor of the union between quantum theory and gravity being a non-computable theory (Geroch and Hartle, 1986). But he also relies to some extent on his own arguments to the effect that consciousness involves an element of non-computability. If this is true (the argument is contentious, but has so far evaded any would-be knock-down blows), then there is no likely place to locate a non-computable process *other* than in the process of reduction. In general, Penrose imagines that the "random effects of the environment would dominate, so OR [objective reduction] would be virtually indistinguishable from the random R procedure that is normally adopted by quantum theorists. However, when the quantum system under consideration remains coherent and well-isolated from its environment, then it becomes possible for its state to collapse spontaneously...and to behave in non-computable rather than random ways." The distinction here between system and environment is somewhat unclear - if the 'system' is *not* well-isolated from its environment, then it would seem that what *was* the 'environment' now becomes part of the 'system'. Nevertheless the argument seems only to require that objective reduction can be 'orchestrated' under certain circumstances that would 'bring out' its non-computable features. This is the line of reasoning that led Penrose, in collaboration with Stuart Hameroff, to formulate a theory of consciousness in which large-scale quantum coherence, originating in the microtubules of neurons in the brain, is 'orchestrated' to a self-collapse that in turn influences brain function (Hameroff and Penrose, 1995, 1996). Of this theory, we shall hear more in later lectures of this course.

Modern Quantum Theory

Quantum field theory

Most physicists do not in fact use quantum mechanics in the general course of their work. In the attempt to better understand nature and to increase the accord between theory and experiment, physicists were led to formulate quantum field theory, which differs in several respects from its predecessor.

Field theories in general (there are classical field theories as well as quantum field theories) result as the number of degrees of freedom in a theory tends to infinity. The objects of ordinary quantum or classical mechanical description are endowed with some finite number of degrees of freedom, their positions and momenta say. In quantum mechanics these may not be precisely defined due to the inherent indefiniteness described by the uncertainty principle, but they nevertheless constitute a definite number of degrees of freedom. In moving to a field description, this number becomes infinite. This infinite number of degrees of freedom corresponds to the fact that the 'field' must take a value at every point in space.

In moving from quantum mechanics to quantum field theory, the position and momentum coordinates are restored to definiteness. In quantum mechanics these coordinates act as locators of an object (in position or momentum space) and so must be, to a degree, uncertain. In field theory, they act only as coordinates and so can be taken as precisely defined. Of course, the conjugate relation that exists between position and momentum in quantum mechanics shows up in quantum field theory, but here it is a relation between

the field itself and its conjugate field, a field derived in a manner analogous to the way one derives momenta in quantum or classical mechanics.

Another distinction of note is the fact that in quantum mechanics, there is only one vacuum, a conclusion derived from the equivalence theorem. By contrast, quantum field theory can take any number of vacua. The relevance of this is that field theory can describe different kinds of dynamics depending on what physical conditions prevail. Taking superconductivity as an example, the same piece of wire might act in an entirely ordinary fashion under a certain set of circumstances, but as these change, may become superconducting. Quantum mechanics is not able to describe this transition since the dynamics in the two different states act out of different vacua.

In part because of the deep connection to the vacuum in field theory, calculations can incorporate the tremendous amount of spontaneous activity that is constantly taking place in vacuum. Using this and a number of other improvements, field theories like quantum electrodynamics (QED) and quantum chromodynamics (QCD) have vastly improved upon the predictions that would be possible if quantum mechanics were the only available tool.

Modern quantum field theory is sometimes touted as the marriage of quantum mechanics to special relativity. Indeed, it does derive principles from both theories, but this should not be taken to imply that there is a seamless union. The tension that exists between quantum mechanics and special relativity may not be so evident in a field theory guise but neither is an explanation of the results of EPR-type experiments.

Quantum field theory may thus have little to contribute towards interpreting quantum theory. While its formulation has allowed some of the most accurate predictions of physical quantities in the history of humankind, its evolution has not significantly elucidated the foundational issues that were raised already with the advent of quantum mechanics. It has however, as a practical tool of the trade, entered into the description of certain quantum theories of consciousness and thus will be encountered from time to time in the course of these lectures.

References

- Albert, D. (1992) *Quantum Mechanics and Experience*. Harvard University Press, Cambridge, Mass.
- Albert, D. and Loewer, B. (1989) Two no-collapse interpretations of quantum mechanics. *Nous* **23**, 169-186.
- Aspect, A., Grangier, P. and Roger, G. (1982) Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: a new violation of Bell's inequalities. *Phys. Rev. Lett.* **49**, 91.
- Aspect, A., Dalibard, J. and Roger, G. (1982) Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.* **49**, 1804.
- Bell, J. S. (1964) On the Einstein-Podolsky-Rosen paradox. *Physics* **1**, 195-200.
- Bell, J. S. (1966) On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **38**, 447-452.
- Bohm, D. and Hiley B. J. (1993) *The Undivided Universe*, Routledge, New York, N.Y.
- Brody, T. A., (1993) *The Philosophy Behind Physics*, L. de la Peña and P. E. Hodgson (eds.) Springer-Verlag, New York, N.Y.
- Caldeira, A. O. and Leggett, A. J. (1983) *Physica* **121A**, 587.
- Chalmers, D. J. (1996) *The Conscious Mind*. Oxford University Press, New York, N.Y.
- Clauser, J., Horn, M., Shimony, A. and Holt, R. (1969) *Phys. Rev. Lett.* **26**, 880-884.
- D'Espagnat, B. (1976) *Conceptual Foundations of Quantum Mechanics*. Benjamin, Reading, Mass.
- DeWitt, B. S. (1970) Quantum mechanics and reality. *Physics Today* **23**, 30-35.
- DeWitt, B. S. (1971) The many-universes interpretation of quantum mechanics. In *Foundations of Quantum Mechanics*, B. d'Espagnat (ed.) Academic, New York, N.Y.

- DeWitt, B. S. and Graham, N. (1973) *The Many-worlds Interpretation of Quantum Mechanics*. Princeton University Press, Princeton, N.J., 155-165.
- Diósi, L. (1989) Models for universal reduction of macroscopic quantum fluctuations. *Phys. Rev.* **A40**, 1165-1174.
- Einstein, A., Podolsky, B. and Rosen, N. (1935) Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777-780.
- Everett, H., (1957) 'Relative state' formulation of quantum mechanics. *Rev. Mod. Phys.* **29**, 454-462.
- Everett, H., (1973) The theory of the universal wave function. In *The Many-worlds Interpretation of Quantum Mechanics*, B. S. DeWitt and N. Graham (eds.) Princeton University Press, Princeton, N.J., 64.
- Gell-Mann, M. and Hartle, J. B. (1989) Quantum mechanics in the light of quantum cosmology. In *Proceedings of the Third International Symposium on the Foundations of Quantum Mechanics*, S. Kobayashi (ed.) Physical Society of Japan, Tokyo.
- Geroch, R. and Hartle, J. B. (1986) Computability and physical theories. *Found. Phys.* **16**, 533.
- Ghirardi, G. C., Rimini, A., and Weber, T. (1980) A general argument against superluminal transmission through the quantum mechanical measurement process. *Lett. Nuovo Cim.* **27**, 293-298.
- Ghirardi, G. C., Rimini, A., and Weber, T. (1986) Unified dynamics for microscopic and macroscopic systems. *Phys. Rev.* **D34**, 470.
- Griffiths, R. (1996) *Phys. Rev.* **A54**, 2759.
- Griffiths, R. (1998) *Phys. Rev.* **A57**, 1604.
- Hameroff, S. R., and Penrose, R., (1995) Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Neural Network World* **5** (5) 793-804.
- Hameroff, S. R., and Penrose, R., (1996) Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. In *Toward a Science of Consciousness - The First Tucson Discussions and Debates*, S. R. Hameroff, A. Kaszniak and A. C. Scott (eds.) MIT Press, Cambridge, MA.
- Hawking, S. W. (1975) Particle creation by black holes. *Commun. Math. Phys.* **43**, 199-220.
- Károlyházy, F. (1966) Gravitation and quantum mechanics of macroscopic bodies. *Nuovo Cim.* **A42**, 390-402.
- Károlyházy, F., Frenkel, A. and Lukács, B. (1986) On the possible role of gravity on the reduction of the wave function. In *Quantum Concepts in Space and Time*, R. Penrose and C. J. Isham (eds.) Oxford University Press, Oxford, U.K.
- Lockwood, M. (1989) *Mind, Brain and the Quantum*. Blackwell, Oxford, U.K.
- Mermin, D. (1997) Non-local character of quantum theory? *Amer. J. Phys.* **66**, 920-924.
- Penrose, R. (1987) Newton, quantum theory and reality. In *300 Years of Gravity*, S. W. Hawking and W. Israel (eds.) Cambridge University Press, Cambridge, U.K.
- Penrose, R. (1989) *The Emperor's New Mind*, Oxford University Press, Oxford, U.K.
- Penrose, R. (1993) Gravity and quantum mechanics. In *General Relativity and Gravitation, Proceedings of the Thirteenth International Conference on General Relativity and Gravitation* held at Cordoba, Argentina 28 June-4 July 1992. Part 1: Plenary Lectures., R. J. Gleiser, C. N. Kozameh and O. M. Moreschi (eds.) Institute of Physics Publications, Bristol.
- Penrose, R. (1994) *Shadows of the Mind*, Oxford University Press, Oxford, U.K.
- Stapp, H. P. (1986) Einstein time and process time. In *Physics and the Ultimate Significance of Time*, D. R. Griffin (ed.) State University Press, New York, N.Y.
- Stapp, H. P. (1997) Non-local character of quantum theory. *Amer. J. Phys.* **64**, 300-304.
- Stapp, H. P. (1998) Meaning of counterfactual statements in quantum physics. *Amer. J. Phys.* **66**, 924-926.
- Stapp, H. P. (1999) Non-locality, counterfactuals and consistent histories, quant-ph/9905055v3.
- Unruh, W. (1997) *Phys. Rev. A*, to appear (quant-ph/9710032).
- Von Neumann, J. (1932) *Mathematische Grundlagen der Quantenmechanik*, Springer-Verlag, Berlin (Engl. tr. by Princeton University Press, Princeton, N.J., 1955)
- Wheeler, J. A. (1957) Assessment of Everett's 'relative state' formulation of quantum theory. *Rev. Mod. Phys.* **29**, 463-465.
- Wheeler, J. A. (1990) Information, physics, quantum: The search for links. In *Complexity, Entropy, and the Physics of Information*, W. Zurek (ed.) Addison-Wesley.
- Zurek, W. in *Non-equilibrium Quantum Statistical Physics*, G. Moore and M. Scully (eds.) Plenum, New York, N.Y.

From: Revue de la Pensee D'aujourd'hui, vol. 24, no. 11, 276-291 (1996)
[in
Japanese translation]. Also appeared in English in Brain &
Consciousness,
Proc. ECPD Workshop, ed. by Lj. Rakic, G. Kostopoulos D. Rakovic, and
Dj.
Koruga, pp. 157-169 (1997)].

Principle of Philosophical Relativity
Michael Conrad
Department of Computer Science
Wayne State University
Detroit, Michigan 48202 USA

Abstract. The internalist and externalist points of view are distinguished by different attitudes about the relationship between what is given in experience, including the feeling of control, and the description of experience in terms of constructs assumed to refer to properties that have an objective existence outside the self. The principle of philosophical relativity asserts that scientific theories should be open to different interpretations on such antinomic issues. Today's major physical theories, quantum mechanics and general relativity, are examined from the point of view of philosophical relativity. A model that incorporates the main but conflicting features of these theories (quantum superposition and gravitational nonlinearity) is briefly outlined, and its significance for understanding the biological life process considered. The analysis supports the idea that the universe as a whole, and biological life in particular, are always chasing an unattainable self-consistency.

Keywords: *Quantum mechanics, General relativity, Fluctuon model, Biological information processing, Consciousness*

1. Internalist versus Externalist Stance

Daily we have the experience of exerting control on the world around us. Daily we also experience limitations on this control. One might take an extreme view and deny either of these facts, or with more caution, admit them as impressions, but assert that they are really very strong illusions rather than facts.

The control point of view we can call an internalist perspective. The no-control point of view we can call the externalist perspective. Many intermediate perspectives are possible, corresponding to the fact that there are aspects of the world over which we can, in practice, exert more or less control. Thus I feel I can control the next series of letters in this sentence, and could have referred to these as a word if I chose. I feel I can change the angle of the monitor I am looking at, and could choose to make it more or less comfortable. I don't feel my thoughts have any influence on the orbit of the moon.

Many scientists by nature, including the present writer, are highly attracted to the external point of view. After all, the form, color and architecture of nature is what natural scientists want to appreciate, describe, and in so far as possible compress into shorter descriptions that we call explanations. One could, as noted above, take this point of view to the extreme and argue that our sense of self and control is an illusion. But the existence

of an experienter is difficult to deny, since it is the precondition of all external experience. The sciences of the mind, such as psychology, have an unavoidable orientation to this direction. Just as students of the natural sciences may find an externalist attitude to be most useful, the student of the mind might find an internalist perspective, with its basis in the phenomenology given in experience, to be most fruitful. One could conceivably pursue along this line, and going to the extreme argue that really the whole world is mind, and that it is the external rather than the internal that is illusory. But the existence of that which cannot be controlled, and the existence of other minds external to us, is at least awkward to deny.

Philosophy has been chasing itself around these two extremes, and covering various intermediate positions, as far back as the records go. The pure externalist attitude has the advantage that it provides a more complete explanation of phenomena that we experience, other than the phenomenon of having experiences, and accordingly has the advantage of affording more control of the world, despite its basic denial of control. The pure internalist attitude has the advantage that it is logically more consistent, since it does not exclude itself, but it has the disadvantage that in practice it provides a much less complete account of experience (since it does not even explain why we can't control the orbit of the moon as easily as we control the motions of our fingers).

One does not have to go to either of these extremes of course. A vast spectrum of intermediate views are possible. To their exponents they all appear sensible. Undoubtedly they all are. The problem is that they share the essential defects of both the pure externalist and pure internalist points of view, with the addition of some arbitrary assumptions designed to hide these defects.

2. Principle of Philosophical Relativity

My goal in the present paper is not to solve the above problem. I regard it as antinomic, in the sense that equally credible arguments can be given to contradictory positions. Issues with this character include freedom versus necessity, the ultimate nature of reality, and whether the universe has a beginning.

What I will do is propose instead that scientific theories should be so constructed that they do not provide definitive answers to inherently arguable issues such as the above, but rather allow for multiple conflicting interpretations. I will call this the principle of philosophical relativity, since it asserts that there are no preferred philosophical coordinate systems. If there is no good reason for choosing between two philosophical stances it is a matter of indifference which one is chosen, except as a personal preference. This does not mean that every philosophical theory is equally good, no matter how well or badly architected, any more than it means that every choice of coordinate system is equally convenient, and that coordinate systems could not be constructed that are of no conceivable use. What it means is that scientific theories should strive as an ideal, possibly an unattainable ideal, to be open to different interpretive perspectives.

Philosophical relativity should not be confused with logical positivism [1]. It does not assert that antinomic questions are meaningless, or mistakes of language, or that answers to them are meaningless. To say that the will is free should not, in short, be analogized to a statement such as "time tastes like a soba noodle". Nor does the principle assert that the only meaningful statements, apart its own statement, are those that are empirically verifiable or mathematically provable. The criterion implied by it is absolutely different.

If scientific theories do not satisfy philosophical relativity then, according to the principle, they should be modified so as to do so. Furthermore, if doing so requires adding a construct to the theory that has no empirical counterpart and is not deducible from an empirically testable assumption of the theory, this is not only permissible but even mandatory. The principle of philosophical relativity is thus a tool of scientific theory construction. Furthermore, it is not a tool that can adhere to entirely positivistic criteria (in so far as these can actually be specified), since it requires theories to include constructs that ensure the facts of experience can be interpreted in antinomic ways. Thus, we may imagine a purely deterministic theory that is in principle complete and consistent. Consistent means that it entails no internal contradictions. Complete means that it could in principle account for all experience (we ignore here the in principle impossibility of ever verifying this). Present such a theory to the man who is inclined to determinism and that man will applaud. For him the theory covers all experience and demonstrates that any feeling of freedom that he possesses is illusory. Present the same theory to the man who believes that he has freedom to choose the next word that he utters, and that man will simply say the theory is not consistent with the facts of his experience. A fortiori, no theory that gives a definitive answer to this issue-resolves the antinomy of freedom versus necessity-can be regarded as demonstrably valid in a public way. The public will always disagree among itself. The pure externalist program may have achieved a victory with the development of our putatively universal deterministic theory, but it is a Pyrrhic victory, for at the very moment of its public presentation the theory will lose the universal objective character that is the ideal of the externalist program.

3. Physics Today

Let us first look at science today, using the philosophical relativity principle as a tool of analysis. It is of course incorrect to speak of science today as if it were a single party platform. There are lots of different theories, models, methods, and approaches, so it would be quite unfair to assert that physics or biology says this or that. It is just individual scientists who say this or that, and accordingly there is plenty of disagreement about what science as a whole says.

Nevertheless, there are two theories-quantum mechanics and general relativity-whose essential ideas have established themselves as having particular foundational value. So let us consider whether these two theories satisfy philosophical relativity.

But first step back for a moment and consider Newtonian mechanics, where the situation is much simpler, since the model is completely deterministic. At best it is possible to create the appearance of indeterminism by considering cases in which the dynamics are highly sensitive to initial conditions. In these cases the future is not computable, but still there is a definitive ontological answer to the question of freedom versus necessity. Furthermore, the description of world in the Newtonian model is completely in terms of real numbers, interpreted as referring to space and time. If the model is regarded as a universal theory it therefore gives a definite answer to the question, "What is the ultimate character of reality?" Reality, it asserts, can be completely described in terms of real quantities. Qualities, such as the quality of red, are only secondary, hence fundamentally illusory. Space and time are also qualities, usually referred to as primary qualities, since secondary qualities such as red are presumably reducible to them. But there is no place for such primary qualities in the theory either, at least as they are given in experience.

There is no place for experience, and therefore for an experiencer that has experience. There is no place then for consciousness, and for the experiencing process that we ordinarily associate with life.

The Newtonian model fails philosophical relativity on all these grounds. It is not only problematic philosophically, it is problematic for biology and psychology as well. It makes it very tough for the biologists to give a unified account of physical and biological phenomena. One can gloss this problem by saying it is all a matter of complexity. But in reality it is a conceptual dissonance that implies a division of the phenomena of nature. In practice much of modern science follows along this line without admitting it. The brain is explained in terms of neurons, regarded as atoms of the nervous system, and neurons are explained in terms of large and small molecules taken as indecomposable entities, and then the chemist comes along and explains something about molecules in terms of electrons and nuclei, and then finally the physicist tries to explain the properties and transmutations of the various elementary particles in terms of concepts that are never called on by the chemist. If one admitted that this is a necessity, at least when it comes to the life process, then one would be making a rather definite statements about monism versus dualism, and about the relation of part and whole, that would in themselves violate philosophical relativity.

Quantum mechanics is a step towards satisfying philosophical relativity, but not a full step. The theory is deterministic so far as the equations of motion are concerned, but introduces randomness in the measurement process. Of course randomness does not by itself mean control or free will. It just means that so far as known the description of some phenomena is not susceptible to compression [2]. But this would also be true of a spontaneous phenomenon, and such a phenomenon would at least not exclude free choice. This is all that is required by the philosophical relativity principle.

Let us look a little more carefully at quantum mechanics, and why it only goes a half step, without here attempting to consider all the multifarious interpretations that have been proposed [cf. 3]. The main feature is the superposition principle. The wave function governing the time evolution of a system is a linear superposition of its possible states. The crucial point is that the possible states interfere with each other. Thus the superposition cannot be a statistical ensemble (unless hidden variables and hence entirely nonclassical actions at a distance are contemplated). An electron in my hand could actually be anywhere, including outside of what appears to be my hand. The possible locations of this single electron in effect interact with each other in a manner reminiscent of the interactions between different actual molecules in a disturbed body of water.

Whenever we think about an electron we locate it somewhere. This is the classical picture, the kind of picture we can have in our conscious experience. Picturing quantum superpositions is an oxymoron (apart from mathematical pictures [cf. 4]). The process of measurement finally means bringing something into our conscious experience in an acceptable way. Thus the superposition must be collapsed into a classical picture, one that can be described by real numbers that ultimately refer to dimensions of space and time. The jump from the set of possible states to an actual state depends on a probabilistic way on weights assigned to each of the possible states This is the process that in today's quantum mechanics is problematically related to the time evolution equations. The latter are reversible and entropy conserving, whereas the measurement process is irreversible and entropy increasing.

The dichotomy between time evolution of the wave function and its collapse in measurement reflects an ambiguous treatment of acceleration. The superposition principle requires the time evolution equations to be linear, otherwise the superpositions would break down. (Here linear means that the changes in the wave function can be described by unitary transformations, i.e., transformations involving rotations in Hilbert space, with no stretching or bending.) So it appears that the quantum mechanical model of acceleration, or change in state of motion, must be linear. Change in state, as measured by the change in the probability distribution of a system, proceeds through the interference of possible states of different energy [cf. 5]. The linear time evolution equations capture this. But actually we can never see that a change in state has occurred until a measurement is made. This is a nonlinear process, since measurement is a nonunitary transformation. When we do look we see that the system has jumped between two of its possible (stationary) states, as when an electron jumps from one energy state to another and either emits or absorbs a photon. In short, the quantum jumps that occur when the measurement process collapses the wave function from possibility space to actuality correspond to the quantum jumps that occur when a system undergoes time evolution. But the former are described by a probabilistic process, whereas the latter are not.

This is the source of the paradoxes of quantum measurement, and of the great controversies over the interpretation of quantum mechanics. No choice is allowed by the equations of motion. But when we make a measurement we make a choice, first as to what measurement is to be made, therefore what aspect of the system is to be made classical, and second as to when to make the measurement. Furthermore, this choice strongly affects the future development of the system, since if we make one aspect (say position of an electron) precise than a conjugate aspect (momentum) must become imprecise. The quantum mechanical model of acceleration allows no choice when it comes to the possible quantum jumps whose occurrence is retrospectively buried in the time evolution equations (the so-called sum over histories), but requires choice whenever a quantum jump is made explicit. The theory thus satisfies philosophical relativity so far as the possibility of choice is concerned, but not in a consistent way.

We can note another feature of the above discussion that on the surface at least appears contradictory. We said that measurement involves the creation of a classical picture in terms of real numbers that ultimately have space-time referents (other quantities, such as mass and charge, are ultimately operationally defined in terms of combinations of space-time measurements). But we also said earlier that space and time are qualities in our experience (sometimes referred to as qualia). We therefore have referred to space and time as something describable in terms of quantities and in the next breath as qualities which should not be obligatorily reducible to quantities, for if the theory imposed such an obligation then it would violate the principle of philosophical relativity. It would do so because it would then entail implications about the ultimate nature of reality, including the implication that the qualityness of our experience is illusory. The pure externalist will applaud this entailment; but the man who takes the qualityness of his experience as a fact of experience would be perfectly entitled to say that any theory which required him to deny this fact did not fit the facts.

The Newtonian model cannot escape this collision with philosophical relativity, since everything is classical. This is not so for quantum mechanics. The unpicturable stratum of

superpositions is not describable by real numbers. Complex quantities are required. If a human observer is a quantum mechanical system, then that observer should be described by a superposition. There is a place for qualities. The superposition must collapse in order for that observer to ever take a definite action. Choice enters. So the theory has qualities (including the qualities of space and time, and many other qualities) as eligible referents, and has decision making (or control with choice) as an eligible referent. The term "eligible" here means that these features are not excluded. It does not mean they are entailed. It is possible to propose that collapse never occurs and that nothing ever becomes definite; or that everything is really definite because hidden variables underlie the superposition. Alternative philosophical interpretations, with antinomic philosophical positions, are possible. This is just what is required of a theory that satisfies philosophical relativity.

But still quantum mechanics is only a half step in this direction, because the observer is not in fact included in the theory. If the observer were included, the measurement process would be embedded in the time evolution process. The theory should at least allow the measurement process to be so embedded, for if it excluded this it would definitely entail dualism. But any such entailment would violate the principle of philosophical relativity. Now let us see how this situation is affected by the other leading idea in today's physics, the general theory of relativity. The central idea is the principle of equivalence. It is impossible to locally distinguish acceleration (or change in state of motion) from a gravitational field. The general theory of relativity is thus a theory of the gravitational field. Furthermore, it culminates in a self-consistent field theory of the gravitational field: mass controls the structure of space-time and the structure of space-time controls the motions of mass [cf. 6]. Thus the theory is inherently nonlinear. Since the equations describing gravitation are nonlinear, and since gravitation is locally equivalent to acceleration, this means that the general relativistic model of acceleration is nonlinear. But this in turn means that the linear superposition principle cannot hold. Superpositions should collapse when gravity is taken into account.

By itself general relativity is an entirely deterministic theory. Accordingly it does not satisfy philosophical relativity. Furthermore, the theory is formulated at an entirely classical level. Operationally speaking it is quite impossible to transform to a coordinate system that eliminates changes in the state of a quantum system, since any such a change in state is an unpredictable jump. Nevertheless, the general point holds: if the macroscopic description of gravity is necessarily nonlinear the micro description should be as well. Superpositions should collapse spontaneously. A theory along this line would satisfy the principle of philosophical relativity, since the time evolution would embed the measurement process within itself. It would not compel anyone to admit the existence of internal observers, or of irreducible choice, or of qualia, but it would not deny anyone the possibility of admitting these features. They would all be eligible as referents of the theory.

4. The Fluctuon Model

At this point I would like to turn to a specific model, called the fluctuon model, which I believe is a step towards a theory that meets the above requirements. Here it must suffice to indicate some of the motivating ideas of the model. Technical treatments can be found elsewhere [7-13]. Nontechnical discussions more extensive than the present one can be found in [14, 15].

At the outset it has to be said that the development of such a theory, one that merges the basic ideas of quantum mechanics with those of relativity, has turned out to be extremely difficult. The problem is that it is difficult to satisfy all the fundamental principles which are generally believed to apply (e.g., symmetries associated with the various conservation laws, the uncertainty principles, microscopic reversibility, second law of thermodynamics). The distinguishing idea of the fluctuon model is that the universe itself cannot satisfy all these fundamental principles simultaneously. We can think of it as a system that keeps changing until it reaches a self-consistent form of organization, at which point change would stop. But final self-consistency (or equilibrium) is never possible. The irremovable inconsistency (or disequilibrium) is the motive power of the time evolution of the universe, both in the narrow sense of the changes in states of motion of particles due to their interactions with one another and in the larger sense of cosmic evolution.

This viewpoint has close conceptual relations to the idea of perpetual disequilibration that has been forwarded by Matsuno [16] and Gunji [17]. Inconsistency and hence generative power, according to Matsuno, is inherent in the inseparability of a system's dynamic development from the boundary conditions that are assumed to confine this development

The physical picture that serves as the starting point of the fluctuon model is a Dirac-like vacuum sea. The so-called vacuum is a sea of unmanifest vacuum fermions. In the Dirac model these are negative energy electrons, but in the eventual development of the fluctuon model the vacuum particles only acquire energy and charge by virtue of being in a superposition of manifest and unmanifest states.

For concreteness consider the electrostatic force between two electrons. The interaction in conventional field theories is due to the exchange of virtual photons. The rough picture is that the electron emits a photon and recoils. Momentum is conserved, but energy is not. The photon is virtual since it can only exist in such a forbidden energy state for a time allowed by the time-energy uncertainty principle. If the photon is absorbed by a second electron its momentum is transferred to that electron.

The exchange process in the fluctuon model is in some respects a generalization of this. The photon is viewed as a virtual electron-positron (or electron-hole) pair, spontaneously manifesting in the vacuum sea due to an energy fluctuon compatible with the time-energy uncertainty principle. The fluctuation energy must at least be equal to the mass-energy of the electron-positron pair. Thus the pair must rapidly decay. If the fluctuation energy is about equal to the mass-energy of the pair it will decay in a reasonably restricted amount of time. Such pair creation is only possible in the neighborhood of a manifest mass, otherwise it would be impossible to conserve momentum in all coordinate systems. We will call the manifest mass (in this case a manifest electron) an absorber. When the pair is transiently excited it recoils from the absorber with a definite quantity of momentum; momentum overall is conserved, as in the conventional picture of a virtual particle exchange. When the pair decays, as it is required to do by the time-energy uncertainty principle, it is no longer next to the absorber. At this point it satisfies conservation of energy, but can no longer satisfy conservation of momentum, since there is no neighboring manifest mass that can assure that momentum is conserved in all inertial coordinate systems. We have thus arrived at our first inconsistency. The only way of resolving it is for the pair to regenerate, and to continue to decay and regenerate until it

collides with a second absorber.

We can think of the virtual photon as a sequence of transient electron-positron pairs that skips through the vacuum somewhat in the fashion of a stone skipping over water, except that there is no friction, due to the quantum nature of the system. In the conventional picture we would only have one skip, and since there would then be no requirement for the length of the skip to match the density of unmanifest vacuum particles (or potentialities for pair production) the skip could have any energy. Furthermore, self-interactions that are excluded in the multi-skip picture would occur. The problems of infinite renormalization that plague conventional field theories derive from these features. The multi-skip picture eliminates these problems altogether.

The model actually comprises three vacuum seas. The gluons that mediate the strong force can be interpreted as chains of transient quark-antiquark pairs. This is the least dense sea (in the fluctuon model the strength of an interaction decreases as the density of the sea increases). The weak force can be interpreted as a variation of the electromagnetic force that occurs due to variations in the density of the electromagnetically active subsea. Mesons mediating attractive interactions can be interpreted as single step exchanges involving quark-antiquark pairs of zero spin that are confined to the neighborhood of the absorbers. The gravitational force is mediated by the supersea of all vacuum fermions. The vast majority of vacuum fermions that contribute to it (called massons in the fluctuon model) do not contribute to the other forces. Thus gravitons are mainly chains of transient masson-antimasson pairs.

The gravitational force is actually an indirect force in the fluctuon model. Manifest absorbers polarize the vacuum sea around them, due to the fact that vacuum particles act as temporary absorbers (being partly in manifest states). The electron subsea is overall homogenous, with only local density depressions surrounding manifest charges, due to the requirement for overall charge neutrality. But mass comes only in a positive charge. Thus the vacuum depressions surrounding masses can become large, and concomitantly the distant elevations in vacuum density can become large. The attractive gravitational interaction is due to masses being pushed together by gravitons emanating from transient vacuum absorbers (called trapped fluctuons) in these distant regions of space. All the masses in the universe accordingly contribute to gravity. The fluctuon model of gravity thus incorporates Mach's principle.

The space curvature of general relativity is identified with the density structure (or curvature) of the full vacuum sea. The electron and quark subseas do not follow this curvature, by virtue of the overall neutrality requirement, and due to their much smaller densities they make a negligible contribution to it. This is why it is possible to make a geometrical model of gravity, and yet have other fields that would seem to require that different geometries cohabit the same universe.

Now let us return to the fluctuon itself. We saw that this is a chain of transient particle-antiparticle pairs that propagates due to the impossibility of simultaneously satisfying all the requisite conservation and uncertainty principles except when it interacts with an absorbing particle. Let us check out the broader consequences of this local inconsistency. The first point is that the existence of the chain is consistent with the uncertainty principle as long as no observer is allowed to ride on top of it. The chain must therefore propagate with the same velocity (light velocity) in all inertial coordinate systems. It is possible to show that a perpetual motion device could be constructed if this were not the

case. The principle of special relativity follows and from this, along with the basic fluctuonic interaction between two electrons (and a few technical assumptions) it is possible to deduce Maxwell's equations [7]. A generalization of the argument leads to the principle of equivalence, and hence to the principle of general relativity. The variations of vacuum density produced by mass and charge mean that fluctuons could not propagate without impermissibly violating conservation of energy unless accompanied by distortions of vacuum density. Electromagnetic and gravitational waves can thus be interpreted in terms of distortions of the vacuum sea.

When absorbing particles change their state of motion in response to the exchange of fluctuons (mainly photons, gluons, gravitons) the density structure of the vacuum sea must change as well. This is most important for the gravitational interaction, since the density structure of the electron and quark subseas is held nearly constant by the overall neutrality requirement. The gravitational forces between absorbers change as a result of this altered density structure. In short, the manifest masses control density curvature, and density curvature controls the motions of manifest masses. This is precisely the same self-consistent field nonlinearity that is the hallmark of general relativity, except that space curvature is replaced by the isomorphic concept of density curvature.

But a new feature enters at this point. The density structure must be consistent with the fluctuation energy required for pair production, therefore with the rest mass of the particle-antiparticle pairs constituting the chain. If the fluctuation energy is too high the fluctuons will not extend over enough of a spatial range to excite a neighboring vacuum particle. If it is too low they will jump over neighboring pairs, yielding an effectively lower vacuum density and hence a stronger interaction (which will itself lead to alterations in vacuum density). As absorbers move and surrounding density structure changes the self-consistent relationship between the distribution of manifest absorbers and the distribution of vacuum particles is disturbed. Self-corrective interactions come into play that move the relationship back towards self-consistency. To the extent that self-consistency is disturbed the forces between absorbers are altered and consequently the motions of the absorbers are altered in a highly random way. The element of randomness concomitant to wave function collapse enters, corresponding to the fact that the nonlinear character of the field equations of gravitation should lead to the collapse of superpositions.

Actually two processes occur. The first is a random alteration (or mutation) in the motions of absorbers in response to the break down of the self-consistent relationship between manifest and unmanifest distributions. This is an entropy increasing process. The second is the self-corrective process that pushes these distributions back to self-consistency. This is an entropy reducing process. In the very high mass, high velocity region both these processes become more pronounced. Superpositional collapse merges into gravitational collapse. But gravitational collapse creates gigantic self-corrective interactions that reverse it. The major crisis of today's physics, the end of the principle of energy conservation in ultimate gravitational collapse, is precluded in the fluctuon model.

5. The Most Powerful Physics Laboratory in the Universe: You

Now we can see why inconsistency is important. If there were no inconsistency in the underlying physics of the universe there would be no self-correction. If there were no self-correction there would be no control. A physics that is perfectly self-consistent, if it could be constructed, would violate the principle of philosophical relativity, since it

would exclude control.

Nonlinearity allows for stability, which is tantamount to control. Quantum mechanics, being essentially linear apart from measurement, does not build the possibility of control into its basic structure. It must be added by augmenting the theory with assumptions that justify statistical mechanics. General relativity, being nonlinear, allows for stability. The microscopic interpretation provided by the fluctuon model enlarges this nonlinearity.

Furthermore, it yields a bona fide error correction system, with a noise generator and basins of attraction. The noise generator is superpositional collapse. The basins of attraction are inherent in the self-consistent field aspect of gravity. The universe envisioned by the fluctuon model is in effect a giant servomechanism. Furthermore, the control capabilities inherent in it extend far beyond those of any classical control system, due to vast parallelism inherent in the superposition principle.

It might be thought, given the apparent precision of today's theories, that control features such as those predicted by the fluctuon model would be far too slight to measure except under the most extreme high mass, high velocity conditions or in some future super-super high energy collider. This is not the case, for two reasons. The first is that the effects are unseen not because they are small, but rather because of their ubiquity in macroscopic matter. The second is that the organism is a much more sensitive measuring instrument than any existing technological measuring apparatus. First consider the ubiquity issue.

Ordinary matter is highly homogeneous as compared to bioorganic materials, which are constituted of a vast variety of molecules in highly choreographed arrangements. The consistency restoring effects of the fluctuon model restrict the possible movements of particles, both in ordinary and biological matter. But the effects are completely randomized in ordinary matter, since there are an enormous number of choreographies (technically called complexions) that are entirely equivalent from the macroscopic point of view. The only consequence that would be discernible would be the stability of macroscopic form, a feature that does not have an entirely clear basis in standard quantum mechanics [18]. This is the feature that is too ubiquitous to be discernible.

The second reason, the extraordinary capacity of biological organisms as detectors, is in a way a restatement of the structural heterogeneity and high dynamic choreography of biological matter. The concept of vertical flow of information (or more generally influence) captures the key feature [14, 19]. Biological organisms are the most powerful instruments for discerning the underlying physics of the universe because of their unrivaled capacity to transduce the macroscopic inputs impinging on them to meso- and microphysical forms, to process them at each of these levels, and then to amplify the these micro processes up to the macro level of action. The term "amplification-transduction cascade" highlights this important aspect of vertical information flow.

Thus a single molecular switching operation at the level of DNA may have dramatically visible macroscopic consequences, such as the differentiation of a biological cell into a skin cell or a liver cell, or the development of an organism into one of two very different forms. Macroscopic environmental signals impinging on the organism are converted to, say, hormonal or nerve signals. The latter are converted to chemical signals within cells, which in turn trigger shape changes in a macromolecule, say a protein, and concomitant molecular actions. The shape changes depend on interactions among the various particles that constitute the protein, including the atomic nuclei (whose positions define the shape) and electrons. The mass of the electrons is small enough for the wave properties of matter

to come to the fore. It is this tight coupling between components heavy enough to have an approximate classical description in terms of shape and components light enough to require a quantum description in terms of waves that allows biological organisms be so much more powerful as detectors than measurement apparatus based on conventional materials. The parallelism inherent in the electronic wave function of the electronic system can thus contribute to the capacity the protein to selectively respond to molecular and physiochemical features in its immediate environment, to make the appropriate shape change, and to take the appropriate action. This is the quantum speedup effect [20]. The consequence, in the case considered, might be to activate a gene, and therefore to release the upward running chain of processes that culminates in the morphological form of the organism. But many other examples could be considered, including molecular level control of the manner in which nerve cells respond to the pattern of impinging inputs [21-24].

Now we can consider how gravity enters the process. The small mass of the electrons makes them susceptible to the restorative processes that maintain the self-consistency between the manifest distribution of matter and its "shadow" in the vacuum. When external inputs impinge on the organism these are transduced to molecular and electronic level events. The atomic nuclei and electrons change their state of motion (or, more accurately, undergo transitions to states of different energy). The arrangement of manifest mass and its vacuum shadow becomes inconsistent. Superpositions collapse. The response of the electrons to the consequent restorative forces influences the motions of the atomic nuclei, leading to the molecular actions (in particular catalytic actions) that culminate in the macroscopic actions of the organism. The tight coupling of atomic nuclei and electrons is extended in this way to a coupling with the vacuum sea. It is an arrangement of matter that feeds on the self-regulatory dynamics inherent in the universe. Furthermore, it is an arrangement that can arise only in mutual co-evolution with the structure of the vacuum sea. One could not simply place molecules in an arrangement corresponding to an organism and expect them to be held together by self-corrective interactions with the vacuum; they would be pulled apart by these interactions unless the vacuum sea was also arranged in a complementary manner. According to the fluctuon model the vacuum shadow is an unseen memory that underlies the fantastic form and coherence of biological organisms.

6. Philosophical Relativity Self-Applied

Have we arrived at model that does enough, and not too much? Doing enough means that it covers the phenomena in a way that allows for antinomic interpretations. Doing too much would mean that it solves the antinomy.

Let us go back to the initial antinomy between the internalist and externalist viewpoints. For the internalist the main fact must be that he has a viewpoint. What is given to him in experience, the qualities that constitute his consciousness, is paramount. We cannot see these in the fluctuon model, since all we can see there are formal symbols or linguistic descriptions. But qualities are eligible referents of these descriptions. since superpositions are not themselves quantities. The internalist would reasonably assert his sense of choice is real, and that he is free to exert control on the world to at least some extent. The fluctuon model allows for this, since it is a control theory. Furthermore, the collapse of superpositions, since it introduces a random element, means that genuine spontaneity is an eligible referent of the model. The internalist might hold that his

experience is private, or at least private for all practical purposes. The transduction-amplification cascades that define the circular flow of influence between the macroscopic level of organism behavior and the excitations of the vacuum sea are so delicate and so history dependent that for all practical purposes it would never be possible to directly experience someone else's pain.

But as noted above, the theory does not compel these interpretations (however attractive they might be to the present author). The pure externalist could still argue that consciousness is just the sum total of the spatio-temporal activities of organisms, and that limitations on the possibility of observing these activities do not have any ontological implications. Or he might argue that even though the theory has random elements there exists a map in principle—a giant table—that we just don't happen to know. It might be that this map has no possible constructive existence, that even writing it down would require more matter than exists in the whole universe. But our determinist-externalist could still adhere to ontological determinism. Or, he might take the tack that someday it should be possible to develop a theory that in all respects is equivalent except that it in principle eliminates superpositions and randomness, presumably at the expense of strange action-at-a-distance interactions. He might even hope that the new theory would make more powerful predictions and that it would therefore demonstrate the nonexistence of freedom and the derivative character of qualia. But this would be going to far, for the theory would then disagree with what the internalist has the right to take as fact. According to the principle of philosophical relativity this new, putatively more predictive theory, would actually be making incorrect predictions. It should be replaced by a theory with constructs that preclude these incorrect entailments; if it is not in principle replaceable in this way it could not lay universal claim to being the better theory.

What would happen if an argument arose as to whether an issue was antinomic? It would be obtuse to argue that the issue of consciousness, for example, has not been the subject of highly competent but conflicting philosophizing. It would be obtuse to argue that the arguments for a flat earth are just as credible as the arguments for a globular earth. But cases could arise in between these extremes that are themselves inherently arguable. To eliminate all such cases one would require a criterion that itself was immune from criticism. This is highly unlikely. As indicated at the outset, philosophical relativity is an attitude and an ideal. It is probably not possible to sharply separate questions that should be definitively answered by a scientific theory from those which should be left open to its interpretation. In fact, in the spirit of the philosophical relativity principle, this itself should be an open question, since clearly the credibility of arguments is a matter of judgment. Recalling the term "inseparability" used by Matsuno [16], we must perhaps in the end think in terms of the ultimate inseparability of scientific description and philosophic interpretation. The inconsistency inherent in this inseparability is quite consistent with our whole model. The impossibility of consistency and of standstill inconsistency is the generator of cosmic evolution in the fluctuon model, and of the evolution of those self-enclosed transduction-amplification circles that we call biological organisms. So also is it the generator of scientific and philosophical advancement.

Acknowledgment. This material is based upon work supported by the National Science Foundation under Grant No. ECS-9409780.

References

- [1] A. J. Ayer, ed., Logical Positivism. (The Free Press, New York, 1959).
- [2] G.J. Chaiten, Gödel's theorem and information, *Int. J. Theoret. Physics* 22 (1982), pp. 941-954.
- [3] J.S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, UK, 1987).
- [4] P.A.M. Dirac, *The Principles of Quantum Mechanics*, 4th ed. (Oxford University Press, Oxford, UK, 1958).
- [5] D. Bohm, *Quantum Theory* (Prentice-Hall, Englewood Cliffs, N.J., 1951).
- [6] C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (W.H. Freeman and Co, New York, 1973).
- [7] M. Conrad, Force, measurement and life, in *Toward a Theory of Models for Living Systems*, J. Casti and A. Karlqvist, eds. (Birkhauser, Boston, 1989), pp. 121-200.
- [8] M. Conrad, Transient excitations of the Dirac vacuum as a mechanism of virtual particle exchange, *Phys. Lett. A* 152 (1991), pp. 245-250.
- [9] M. Conrad, The fluctuon model of force, life, and computation: a constructive analysis, *Appl. Math. and Computation* 56 (1993a), pp. 203-259.
- [10] M. Conrad, Fluctuons-I. Operational analysis, *Chaos, Solitons & Fractals* 3 (1993b), pp. 411-424.
- [11] M. Conrad, Fluctuons-II. Electromagnetism. *Chaos, Solitons & Fractals* 3 (1993c), pp. 563-573.
- [12] M. Conrad, Anti-entropy and the origin of initial conditions, *Chaos, Solitons & Fractals* 7 (1996a), pp. 725-745.
- [13] M. Conrad, Fluctuons-III. Gravity. *Chaos, Solitons & Fractals* (1996b), pp. 1261-1303.
- [14] M. Conrad, Cross-scale information processing in evolution, development and intelligence, *Biosystems* 38 (1996c), pp. 97-109.
- [15] M. Conrad, Percolation and collapse of quantum parallelism: a model of qualia and choice. In *Toward a Science of Consciousness*, S.R. Hameroff, A.W. Kaszniak, and A. C. Scott, eds. (The MIT Press, Cambridge, MA, 1996d), pp. 469-492.
- [16] K. Matsuno, 1989, *Protobiology: Physical Basis of Biolog.* (CRC Press, Boca Raton, FL, 1989).
- [17] Y.-P. Gunji, Global logic resulting from disequilibrium process, *BioSystems* 35 (1995), pp. 33-62.
- [18] R. Penrose, R. The Emperor's New Mind, (Penguin, New York, 1989).
- [19] M. Conrad, Molecular computing. In *Advances in Computers*, M.C. Yovits, ed. (Academic Press, Boston, 1990), pp. 235-324.
- [20] M. Conrad, Quantum molecular computing: the self-assembly model, *Int. J. Quant. Chem.: Quantum Biology Symp.* 19 (1992), pp. 125-143.
- [21] G. Matsumoto and H. Sakai, Microtubules inside the plasma membrane of squid giant axons and their possible physiological function, *J. Membrane Biol.* 50 (1979), pp. 1-14.
- [22] E.A. Liberman, S.V. Minina, N.E. Shklovsky-Kordy, and M. Conrad, Changes of mechanical parameters as a possible means for information processing by the neuron (in Russian) *Biofizika* 27 (1982), pp. 863-870 [Translated to English in *Biophysics* 27 (1982), pp. 906-915].

[23] E.A. Liberman, S.V. Minina, O.L. Mjakotina, N.E. Shklovsky-Kordy and M. Conrad, M., Neuron generator potentials evoked by intracellular injection of cyclic nucleotides and mechanics.

Consciousness in David Bohm's ontology

by Paavo Pylkkänen
Consciousness Studies Programme
Department of Humanities
University of Skövde
P.O. Box 408
S-541 28 Skövde
Sweden
E-mail: paavo@ihu.his.se

Introduction

David Bohm was one of the foremost physicists of his generation and made significant contributions to e.g. plasma physics and the foundations of the quantum theory. As is well known, his interests and explorations extended far beyond physics. His biography *Infinite Potential* by David Peat, for example, makes it clear that a very primary interest in Bohm's life was with the question of how to bring about a good society. His engagement with physics, philosophy, social theory, religion and eventually a form of group dialogue all were strongly connected to his concern for "making the world a better place". It is fair to say that Bohm was most concerned about consciousness in the sense of trying to bring about a kind of mind and way of being that would contribute to an overall coherence and creative harmony for humanity. This is not to deny that a great deal of Bohm's work can be seen to be motivated by a more intrinsic interest, in the sense that the issues that are explored are taken to be interesting in themselves, without clear connections to the broader social aims. But it is important to appreciate just how central for Bohm was the concern with transforming society through transforming both individual and social consciousness. In this course our topic is "quantum approaches to consciousness", and on this I will now focus. However, references to Bohm's work concerned with the broader ideas are also made from time to time in the text below.

Our question in this lecture is: "What were David Bohm's ideas about the relationship between quantum theory and consciousness?" This is no easy question to answer, for Bohm's views about the quantum theory developed and changed, as did his views about the mind and consciousness.

I have written some new material for this lecture and also included some previously published material, as well as references to websites. There is in particular

- 1) Historical overview of Bohm's work on mind and matter , specifically written for this course (by Paavo Pylkkänen)
- 2) An exposition and analysis of Bohm's early (1951) analogies between quantum processes and thought, specifically written for this course (by Paavo Pylkkänen).
- 3) Web reference to Bohm's 1990 article "A New Theory of the Relation between Mind and Matter", originally published in the journal *Philosophical Psychology*. In this paper Bohm summarizes his quantum approach to mind and matter.
- 4) Web reference to the physicist Chris Dewdney's home page where you can find a good summary and visual illustrations of Bohm's "causal interpretation" of the quantum theory.
- 5) Introductory material to Bohm's quantum approach to mind, as well as a discussion of it (by Paavo Pylkkänen, modified for this course).
- 6) Web reference to Basil Hiley's overheads in the Tucson III conference, which make an overall presentation of the field of quantum approaches to consciousness, and explain some of his latest ideas on the topic.
- 7) A brief paper by Paavo Pylkkänen on Bohm's view of causality as it develops in the early 1960s in the *Bohm-Biederman correspondence* (Routledge, 1999).

1. Historical overview of Bohm's work on mind and matter

The idea of this section is to briefly review some of Bohm's major contributions relevant to "quantum approaches to consciousness". The following list is not meant to be exhaustive, but tries to give a basic orientation. It expands more on Bohm's lesser known work and is relatively brief in the well known work (e.g. the implicate order).

* 1940s -> groundbreaking contributions in some more standard areas of physics like plasma physics. Bohm suggested that the plasma behaved in some respects like a living organism. If you disturbed it it would shield itself etc.

* one of the clearest formulations of the "standard" interpretation of the quantum theory in his 1951 text book *Quantum theory*. Particular emphasis on making explicit the physical meaning of the theory (as opposed to a mere exposition of the mathematical formalism). Yet Bohm said that after writing the book he still felt he couldn't understand the theory well enough. The book also has a well known section where analogies between quantum processes and inner thought processes are discussed (to be considered in section 2 of this lecture).

* dissatisfaction with standard quantum theory catalyzed by discussions with Einstein led Bohm to "do the impossible" (Bell) - to offer a consistent alternative interpretation of the quantum theory, published in 1952 as two papers in the journal *Physical Review*. According to this interpretation a quantum system like an electron always a particle accompanied by a new type of field. This made possible an unambiguous ontology where particles were moving under the action of not only classical potentials but also a new quantum potential. A striking non-classical feature was non-locality or action at-a-distance, a feature of quantum mechanics which Einstein, Podolsky and Rosen had made explicit in their 1935 attempt to argue against Bohr's conventional quantum theory. Because of the uncertainty principle it is not possible to verify Bohm's interpretation, and it remains partly a hypothesis about what might be going on where we are not able to look. At the same time it provides perhaps the most elegant account of the observed facts of the quantum theory, at least for those feel theories are allowed or even supposed to make assumptions about the unobservable to account for unintelligible observed facts.

Bohm called his approach initially a "hidden variable" interpretation, and a bit later he called it the "causal" interpretation. A yet another name used is "pilot-wave" interpretation. Actually de Broglie had proposed something very similar to Bohm's theory already in the 1920s under the title "theory of a double solution", but soon dropped the idea due to heavy criticisms especially from Wolfgang Pauli. Bohm was able to answer these and as a result de Broglie returned to develop his original approach. Thus it is common to refer to the view also as "de Broglie - Bohm interpretation". There one needs to remember that regardless of striking similarities there are also important differences. One of these is that de Broglie's view sees the electron as a singularity of a field, whereas Bohm's 1952 papers postulate the electron to have a particle aspect which exists over and above the wave aspect.

Bohm develops his approach in co-operation with e.g. Vigier and later on with Basil Hiley and their research students like Chris Dewdney at Birkbeck College, University of London. More description, references and "movies" of Bohm's approach, as well as a description of recent developments, can be found at Chris Dewdney's home page and you are recommended to check this out later on, in section 4 of the lecture at:

<http://www.phys.port.ac.uk/fpweb/index.htm>

For an exposition of a more "mechanistic" line of development (known as "Bohmian mechanics") that has arisen from Bohm's work, see Sheldon Goldstein's home page.

The importance of Bohm's 1952 approach to the "quantum consciousness" debate arises especially through a later re-interpretation of the model, where Bohm suggested that the wave aspect of the electron contains "active information" that in a subtle "non-mechanistic" way in-forms (rather than mechanically pushes and pulls) the particle aspect. In this week's lectures this idea will be extensively discussed, as follows. In section 3 of this lecture it is recommended that you read Bohm's own paper on the topic at

<http://members.aol.com/Mszlajak/BOHM.html>

Paavo Pylkkänen's introduction into the approach and a discussion of Bohm's above paper is also included later on in this lecture, in section 5. Section 6 of the lecture then suggests that you go through Basil Hiley's overheads to the Tucson III meeting at

<http://www.bbk.ac.uk/tpru/welcome.html>

Let us go back to the historical overview. The next philosophically important step is

* the book *Causality and Chance in Modern Physics* (1957, republished by Routledge in 1984), where the causal interpretation is placed into a wider, non-mechanistic world-view, where "qualitative infinity of nature" is a key concept. The critics had accused Bohm for making a conservative step back towards the mechanical Newtonian deterministic physics in his 1952 "hidden variable" papers, and Bohm takes great pains to deny that his causal interpretation implies that nature is a deterministic machine. In a dialectical fashion he sees determinism and indeterminism as two sides of any process in nature. In some domains one of these sides, say, determinism may dominate so that a deterministic theory provides a correct account of that domain (e.g. in the domain of classical physics, there seems to be no principal limit to the accuracy in

which we cannot predict the behaviour of an individual system). Yet if one were able to see that process from "another side" or more precisely one would see indeterminate features coming in (e.g. the more accurate study of particles like electrons showed basic limitations in our ability to predict their behaviour, thus revealing the aspect of indeterminism). Bohm's controversial idea then was that in an even more accurate study, revealing a "sub-quantum level" we might well discover a new kind of determinism, and his 1952 causal interpretation showed one such possible deterministic model. But the key point was that even if some such model were found to be correct, it would not prove that nature is absolutely deterministic, for Bohm's scheme suggests that we would be likely to find a new kind of indeterminacy even at the sub-quantum level in an accurate enough study.

He emphasized that each theory in science is correct only with respect to a particular domain and that similarly, determinism and indeterminism are features of some limited domain (such as the domain of classical physics, quantum physics, subquantum etc.) The essence of "mechanistic philosophy" was defined as "...the assumption that the great diversity of things that appear in all of our experience, every day as well as scientific, can all be reduced completely and perfectly to nothing more than consequences of the operation of an absolute and final set of purely quantitative laws determining the behaviour of a few kinds of basic entities or variables. In this connection it must be stressed, however, that the mere use of a purely quantitative theory does not by itself imply a mechanistic point of view, as long as one admits that such a theory may be incomplete. Hence, mechanism cannot be a characteristic of any theory, but rather ... a philosophical attitude towards that theory" (pp. 38-9).

* in a correspondence with the American artist Charles Biederman he further develops his views about necessity and contingency, and begins to consider the concepts of order and structure as the fundamental concepts in the description of nature (Bohm and Biederman 1999). This reflects itself in 1965 papers such as "Space-Time considered as Discrete Structural Process" and "Problems in the Basic Concepts of Physics", in a paper on the notion of "order" in biology (1969) and eventually in two papers on the notion of the implicate order 1971 and 1973, which were republished in *Wholeness and the Implicate Order* (1980).

For a brief description of how necessity and contingency are understood in the BB correspondence, see my paper "Quantum interweavings", included as section 7 of this lecture.

* a paper discussing "The Problem of Truth and Understanding in Science" appears in Popper's *Festschrift* in 1964.

* discussions with the Indian-born teacher J. Krishnamurti begin in the early 1960s and continue till Krishnamurti's death in 1986. This offers Bohm, among other things, a "method" of introspection; consciousness and the possibilities of its transformation become a central focus of his work. Joint books are published, in particular *The Ending of Time* and *The Limits of Thought* (posthumously). The interest in Krishnamurti combines with Bohm's earlier concerns with the social dimension, and in the 1980s we see Bohm advocating a form of large group dialogue as means of social exploration, transformation and well-being.

* the 1965 text book *The Theory of Special Relativity* contains an appendix "Physics and Perception" which argues that Relativity describes human perceptual process better than Classical Physics. On the whole during this period we see in Bohm's writing more concern with trying to understand human mental processes as a special and developed case of general processes of "reflection" to be found in nature. The later discussions of consciousness in the context of the notion of implicate order bear a clear resemblance to these ideas in the early sixties, both in the *Bohm-Biederman correspondence*, the article "Space-Time considered as Discrete Structural Process" and in the appendix "Physics and Perception".

* the 1973 paper "Human nature as the product of our mental models" considers the idea that our mental reality is in some sense produced by our mental models. This anticipates the later (1980s) ideas, where meaning is seen as the key factor of being. A number of other philosophical and psychological papers appear in the mid 1970s.

* the 1977 paper "Science as Perception-Communication" summarizes Bohm's study of Bohr's interpretation of the quantum theory, where communication plays a central role. The book *Fragmentation and Wholeness* shows Bohm's concern with the mechanization of modern society. A new more holistic and process-like mode of language, the Rheomode, inspired by quantum theory is introduced. Another paper explores the implications of the idea that both reality and knowledge are considered to be processes.

* the book *Wholeness and the Implicate Order* of 1980 presents together six previously published papers and introduces a new one, titled "The enfolding-unfolding universe and consciousness". The *Revision*

magazine features Bohm "holographic universe" alongside Pribram's holographic theory of the brain, and Renee Weber's interviews convey Bohm's approach to a wider public.

- * the book *Unfolding Meaning* (1985) (edited by Donald Factor) includes the paper "Soma-significance" where meaning is seen as a key factor of being. The ideas about dialogue in the sense of a special form of group process begin to take concrete shape. The book *Changing Consciousness* with Mark Edwards presents Bohm's approach to the human condition. Later publications along similar lines include the posthumously published book *On Dialogue*.

- * the anthologies *Beyond Mechanism* (ed. by D. Schindler) and *Physics and the Ultimate Significance of Time: Bohm, Prigogine and Process Philosophy* include important articles by and on Bohm.

- * the paper "A new theory of the relation between mind and matter" (1986) presents the idea of active information for the first time.

- * The book *Science, order and creativity* (1987) with David Peat presents a popular account of many of Bohm's ideas.

- * Bohm's festschrift *Quantum Implications* (edited by Hiley and Peat) comes out 1987. In his own paper Bohm explains the relationship between the hidden variables and the implicate order. A number of other papers by eminent authors discuss issues relevant to "quantum consciousness."

- * in the paper "Meaning and information" in the anthology *The Search for Meaning* (1989) he further develops the ideas of soma-significance and active information. A developed version of the paper "A New Theory of the relationship between mind and matter" appears in the journal *Philosophical Psychology* in 1990. See <http://members.aol.com/Mszlajak/BOHM.html>

(this paper is also recommended reading for this lecture, see section 3 below)

- * Bohm has discussions with the biologist Rupert Sheldrake, where interesting connections are explored between "formative fields" in physics and biology. These are published in the 2nd edition of Sheldrake's book *The New Science of Life*.

- * Bohm reconsiders his ideas on cognition in the late 1980s and early 1990s in co-operation with Paavo Pylkkänen, who is writing a PhD on Bohm and cognitive science. The co-operation results in a manuscript of an article, which is not completed due to Bohm's death 1992. It may be published posthumously in some form.

- * Completed before Bohm's death in 1992, *The Undivided Universe* with Basil Hiley (Routledge, 1993) summarizes the work on the ontological interpretation and explores new developments through the notion of implicate order, also discussing the mind-matter relationship. The book is hailed by C.J. Isham as "One of the most important works on quantum theory during the last twenty years".

- * David Peat's biography of Bohm, *Infinite Potential* appears in 1994.

- * posthumous publications include *On Creativity* (edited by Lee Nichol, Routledge), which includes many articles on mind. The first volume of the *Bohm-Biederman correspondence* (ed. by P. Pylkkänen) is published by Routledge in 1999. Bohm's widow, Mrs Saral Bohm has played a key role in helping this as well as other posthumous publications to appear.

As already mentioned, one of the difficulties of stating Bohm's quantum approach to consciousness is that he explored many different views of quantum theory, opening up different ways to discuss consciousness in a quantum context. It is easy to confuse these different ideas with each other. Before discussing the 1951 analogies, let me therefore briefly summarize how Bohm's mind-matter ideas from different periods relate to the different views of quantum theory that he explored.

Phase 1: the conventional interpretation of the quantum theory

- * until 1951 Bohm was the supporter of the "conventional" interpretation of the quantum theory, Bohr's approach in particular. Thus his 1951 discussion of analogies between thought and quantum processes are based on the conventional approach and must not be confused with those of his later (1980s) ideas on mind and matter (e.g. active information) that arise from a re-interpretation of his own 1952 causal or ontological interpretation.

Phase 2: hidden variables and the qualitative infinity of nature

- * the 1952 papers on hidden variables do not explicitly discuss the mind-body problem. But they are philosophically relevant because they present a "realist" alternative to the prevailing "positivist" Copenhagen view. That is, Bohm argues that unlike Bohr had claimed, it is both possible and meaningful to think about quantum processes at the level that we cannot observe. This is a typical tension between realism and positivism, and Bohm explicitly mentions positivism as his opponent in the papers. The philosophical point is that Bohm tries to show that there exists a mind-independent well-defined reality at

the quantum level in the sense that particles, for example have a well-defined position and momentum. The quantum world exists ontologically independently of the human mind, and the quantum world gives rise to the classical world in situations where the effect of the quantum potential is negligible. In contrast Bohr and others had claimed that quantum theory gives us an epistemological lesson, in the sense that we are not justified in assuming the existence of a well-defined physical reality at the quantum level, independently of the observational phenomena that are known to us; and those observational phenomena do not reveal a well-defined quantum world, but only certain features (like position) at one context, while others equally important (like momentum) are unobservable (complementarity).

One of the characteristic features of philosophy since Kant has been to emphasize the role of the human mind in constructing our experience of the world and to be silent about the nature of the unobserved and unobservable world that is not constructed by us. In this general sense Bohr clearly belongs to the Kantian tradition, as did the positivist (though see e.g. Plotnitsky's book *Complementarity* for considerations that Bohr also radically breaks away from the Kantian tradition). Scientific realists, on the other hand, oppose Kant and the positivists and claim that it is possible to know the "real world", at least to a certain approximation (truthlikeness or verisimilitude). Bohm is clearly a representative of scientific realism in his 1952 papers.

Commentators like Stapp have been keen to note that Bohm's 1952 approach leaves out the mind of the observer from the physical universe. However, this would only follow if Bohm claimed that his version of the quantum theory is a final and complete description of the physical universe, and this he didn't claim or even imply. But perhaps the point was not made clear enough in the 1952 papers, given that so many people took Bohm to represent simplistic determinism and even a kind of elimination of consciousness from the scientific view of the world.

* in any case already in the 1957 book *Causality and Chance in Modern Physics* Bohm explicitly professes the view that no theory can ever plausibly be considered complete, due to his assumption that the universe is infinite not only in a quantitative but also in a qualitative sense. He clearly makes room for the existence of consciousness in the physical world, although he doesn't discuss what consciousness is and how it relates to the physical world.

Phase 3: Back to the conventional interpretation of qm and towards the Implicate Order

* partly as a result of heavy criticisms of the 1952 causal interpretation and partly as a result of not seeing how to develop the approach further, Bohm seems to give up the causal interpretation for a number of years, and starts to reconsider Bohr's approach, as well as to consider a wider framework in which the different contradictory theories of physics (especially quantum theory and general relativity) could be reconciled. This leads later on to the idea of the implicate order. It is very important to realize that the ideas of the implicate order are in some respects more similar to Bohr's approach than to Bohm's own 1952 approach. For example, Bohm begins to consider in 1960s that we must give up the traditional notion of space-time. In the early papers from mid-1960s one finds references to Bohr's discussion of the indivisibility of the quantum of action as important in this context but no references to Bohm's own 1952 work. Bohm's concept of mind as a highly developed process of "reflection" (in some ways similar to "reflection" occurring at the level of physics, with light etc.) begins to emerge in those papers in the mid-1960s, but the 1952 causal interpretation does not seem to play any significant role in that period. Thus one might say that the sort of "quantum approach to consciousness" that Bohm presents in the final chapter of the book *Wholeness and the Implicate Order* is in some ways more in the line with Bohr's approach than it is in the line of his own 1952 approach.

phase 4: the grand synthesis: active information and the implicate order

* Bohm started to reconsider the 1952 work in the late 1970s and it is only in the 1980s that we see Bohm developing a "quantum approach to mind" in relation to the re-interpreted 1952 causal interpretation, with the suggestion that the quantum field contains "active information" which guides the particle aspect. To understand how Bohm himself looked at the relation between the 1952 approach and the implicate order, it is worth reading his own paper in his Festschrift *Quantum Implications*. The idea is that eventually one could look at "active information" as an example of the implicate order, and that both concepts are necessary to work out the Bohmian approach to mind and matter. But given that Bohm's views changed a lot, it is no easy task to spell out the "grand synthesis". Indeed, it seems that Bohm felt that these ideas were very much in the early stages of development.

With these remarks in mind, let us now move on to consider Bohm's 1951 ideas about analogies between quantum processes and thought.

2) Analogies between thought and quantum processes (1951)

Probably Bohm's first published remark on the relation between quantum theory and thought processes can be found in the text book *Quantum theory* (1951, pp. 168-172). Here is a summary and a brief discussion of those ideas.

2.1 Three different analogies

a) *Effects of observation*

"If a person tries to observe what he is thinking about at the very moment that he is reflecting on a particular subject, it is generally agreed that he introduces unpredictable and uncontrollable changes in the way his thought proceeds thereafter."

This is similar to the effects of observation at the quantum level. In other words, the way observation of thought introduces unpredictable and uncontrollable changes in the way thought proceeds thereafter.

is analogous to

the way observation of the position of a particle introduces unpredictable and uncontrollable changes in the particle's momentum.

For example, let us ask a person to give a detailed description of what he is thinking about while reflecting on some definite subject. The "paradox" is that as soon as he begins to give this detailed description, he is no longer thinking about the subject in question, but is instead thinking about giving a detailed description. The idea is that the thought process does not allow itself to be observed beyond a certain approximate level without changing in a significant way.

The analogy with quantum level observation could be a mere coincidence. But one alternative is that the physical aspect of thought involves quantum processes in some crucial way. This would explain in a qualitative way why the direction ("momentum") of thought is disturbed by an attempt to define the content ("position") of thought. This applies especially in the debate about the interpretation of the quantum theory, for there it is widely held that as you define your position, you lose your momentum... :-)

b) *Relationality of being (indivisibility)*

"...if a person attempts to apply to his thinking more and more precisely defined elements, he eventually reaches a stage where further analysis cannot even be given a meaning". This leads Bohm to suggest that A part of the significance of each element of thought processes originates in its indivisible and incompletely controllable connections with other elements.

which is analogous to

The way some of the essential properties of a quantum system (e.g. whether it is a wave or a particle) depend on indivisible and incompletely controllable connections with surrounding objects.

Similarly, "...part of the connotation of a word depends on the words it is associated with, and in a way that is not, in practise, completely predictable or controllable (especially in speech). In fact the analysis of language, as actually used, into distinct elements with precisely defined relations between them is probably impossible."

Comment: Again, this analogy could just be a coincidence. But if the physical aspect of thought and language were essentially involving quantum processes, some holistic features of language and meaning would get a qualitative naturalistic explanation.

c) *Logic and causal laws, concepts and objects, nonlogical discovery and quantum jump*

The thought process is analogous to the classical limit of the quantum theory.

"The logical process corresponds to the most general type of thought process as the classical limit corresponds to the most general quantum process."

The rules of logic are analogous to the causal laws of classical physics.

"Logically definable concepts play the same fundamental role in abstract and precise thinking as do separable objects and phenomena in our customary description of the world."

"Yet, the basic thinking process probably cannot be described as logical."

Sudden emergence of a new idea is analogous to a quantum jump.

Comment: The idea here is that the structure and operation of the human mind reflects the dual nature of the reality out of which the mind emerges. According to Bohm the classical world of separable objects and phenomena governed approximately by the causal laws of classical physics is a real but not the most fundamental aspect of the physical world. The more fundamental level is the quantum level where features like quantum jumps are central. And just as the physical world at the classical limit can be thought to consist of separable objects and phenomena governed by classical laws, so the human mind has a "classical level", a level of logical thought process which consists of logically definable concepts governed by the

rules of logic. Yet this is not the full story about the physical world nor about the human mind. In the physical world at the quantum level, the classical notion of separable objects is not generally applicable, neither are causal laws of the classical type. Instead the essential properties of a quantum system (e.g. whether it is a wave or a particle) depend on indivisible and incompletely controllable connections with surrounding objects. The traditional idea of a separable object thus breaks down at the quantum level. Also, processes are radically different from processes at the classical level (e.g. because of quantum jumps). Similarly for Bohm the "most general type of thought process" cannot be described as logical, and it would be natural to suggest that logically definable concepts are not applicable to the basic thinking process. So, just as there are, if you like, two physical worlds (the general quantum world and the special case of a classical world) so there are two minds (the mind in the sense of a general alogical and aconceptual thinking process and the special case of the mind as logical thinking process with logically definable concepts).

We have the quantum world of inseparable objects and discontinuous processes, and the classical world of separable objects and causal, continuous processes.

We have the aconceptual mind (cf. Pylkkö, *The Aconceptual Mind*, John Benjamins 1998) with alogical processes and the conceptual mind engaged in logical thinking as a special case of it.

Again, we have a strong analogy, and one explanation of it would be provided if it is the case that the physical aspect of the alogical, aconceptual thought process involved quantum processes (with inseparability and discontinuity), while the physical aspect of the logical and conceptual thought process involved classical processes (e.g. classically describable neural "activation patterns" governed by the classical laws of physics).

This last analogy is perhaps the philosophically most sophisticated one, for recent analyzes in the philosophy of mind and of cognitive science have shown the importance of the non-conceptual or aconceptual level of the human mind. The interesting question is what might be the physical or computational concomitants of non-conceptual processes, and connectionism has been pointed to as a candidate. Yet there are arguments that connectionist models are mechanically computable and thus deterministic which would make them an implausible candidate for the physical aspect of a truly non-mechanical level of aconceptual mental processes.

2. 2 Is the analogy between quantum processes and inner experiences / thought processes a mere coincidence?

"Bohr suggests that thought involves such small amounts of energy that quantum-theoretical limitations play an essential role in determining its character."

Much of the enormous amount of mechanism in the brain must probably be regarded as operating on a classically describable level. But "...Bohr's suggestion involves the idea that certain key points controlling this mechanism (which are, in turn, affected by the actions of this mechanism) are so sensitive and delicately balanced that they must be described in an essentially quantum-mechanical way."

If Bohr's hypothesis could be verified, it would explain in a natural way a great many features of our thinking.

"If it should be true that the thought processes depend critically on quantum-mechanical elements in the brain, then we could say that thought processes provide the same kind of direct experience of the effects of quantum theory that muscular forces provide for classical theory."

"We suggest that ... the behavior of our thought processes may perhaps reflect in an indirect way some of the quantum-mechanical aspects of the matter of which we are composed."

Bohm might have added that logical thought processes reflect and perhaps give us some experience of processes in the classical domain, given that he emphasized the analogy between these two in the previous section.

Note that in these remarks Bohm is not directly addressing the "hard problem" of consciousness but is rather focussing on the nature of the thought process. The first point concerns how the thought process behaves when it is "observed"; the second concerns the underlying holistic character of the thought process; and the third notes that there is both a "general" and a "classical" aspect to thought, just as there is a general and a classical aspect to physical processes.

Given that thought processes and quantum processes have so much in common, then perhaps this is so because in some respects they are the same process.

Note that the critics of "quantum consciousness" sometimes claim that people in the approach think "consciousness is a mystery, quantum mechanics is a mystery, perhaps they are the same mystery?". But

even Bohm's 1951 remarks argue that quantum processes and mental processes are not only mysterious but share certain basic properties in common, so much so that we are led to think that they to some extent at least are the same process.

3) Bohm's 1990 article "A New Theory of the Relation between Mind and Matter"

It is now recommended that you read this article. It was originally published in the journal *Philosophical Psychology*. In this paper Bohm summarizes his quantum approach to mind and matter.

[<http://members.aol.com/Mszlajak/BOHM.html>](http://members.aol.com/Mszlajak/BOHM.html)

4) Chris Dewdney's summary and visual illustrations of the Bohm interpretation of the quantum theory.

It is now recommended that you check out Dewdney's website at

[<http://www.phys.port.ac.uk/fpweb/index.htm>](http://www.phys.port.ac.uk/fpweb/index.htm)

5) More introductory material to Bohm's quantum approach to mind by Paavo Pylkkänen

This material is useful to you especially if you have little background in the topic. The text discusses in particular Bohm's 1990 article (section 3) but includes a broader philosophical perspective as well.

**will be added shortly **

6) Basil Hiley's overheads in the Tucson III conference

Here is an overall presentation of the field of quantum approaches to consciousness, and an explanation of some of Hiley's latest ideas on the topic. It is recommended that you first read the "Pre-conference workshop slides" and then the plenary talk slides.

[<http://www.bbk.ac.uk/tpru/welcome.html>](http://www.bbk.ac.uk/tpru/welcome.html)

7) "Quantum interweavings" by Paavo Pylkkänen

This is a brief paper on Bohm's view of causality as it develops in the early 1960s in the *Bohm-Biederman correspondence* (Routledge, 1999). The paper was originally written for Dr. Ivan Havel's 60th birthday volume (to be published) in 1998.

QUANTUM INTERWEAVINGS

Paavo Pylkkänen
Department of Humanities
University of Skövde
P.O. Box 408
S-541 28 Skövde
Sweden
E-mail: paavo@ihu.his.se

Ivan Havel and I first met in 1991. We shared an interest in David Bohm's work and what Ivan likes to call a "transdisciplinary " approach. My own graduate work dealt with applying quantum theory to the mind-matter problem and the like and I was very encouraged by Ivan's open and supportive attitude - after all many researchers still think you are being non- rather than trans- or inter-disciplinary when you want to work across and beyond the boundaries of established disciplines. Over the years we have managed - often jointly - to organize a number of small meetings and workshops to nurture the transdisciplinary conspiracy. I'd like to congratulate Ivan on his birthday by briefly discussing one of the big questions of this century, that of determinism vs. indeterminism in quantum theory, and by evaluating Bohm's role in this debate. Towards the end of the paper I will take a transdisciplinary turn...

1. Indeterminism vs. determinism in quantum physics

The determinism of Newtonian physics was an important cornerstone of the

mechanistic world picture, and for many the most radical feature of the quantum revolution in physics was the failure to predict the behaviour of individual quantum systems such as electrons. From this failure to predict it was generally inferred that such systems have genuinely indeterministic features. Of course, there is much about, say, electrons and atoms that appears as straightforwardly determined, properties such as mass or charge of the electron or energy levels of the atom. Further, in a two-slit experiment a sufficiently large number of electrons produce a precisely predictable interference pattern. And even the behaviour of an individual electron is determined in the sense that we can predict what it will not do: for example it will not go to certain areas of the photographic plate in a two-slit experiment. Yet the fact that we cannot predict just where in the possible areas it does appear in the plate is an example of the radical kind of unpredictability that is also behind the more generalized notion that individual processes at the most basic level of nature known to us are genuinely indeterministic.

It is against this background that one needs to understand the significance of Bohm's 1952 papers on hidden variables, presenting a "causal interpretation of quantum theory" for a first time in a consistent way. Bohm did not, of course, find a way of predicting where, say, an electron will appear in a typical interference experiment. Nor has anyone since been able to do that - quantum unpredictability and the uncertainty principle still reign regardless of important advances in technology (e.g. advances in observing individual atoms with lasers). What Bohm did was to produce a model which showed that regardless of our inability to predict it is still a coherent and physically consistent possibility that the behaviour of electrons actually is determined. Here one must remember those who, like von Neumann, had presented proofs that such models are simply inconsistent with the quantum formalism, a formalism that had gained tremendous experimental support. What Bohm established was that it is ultimately a matter of faith whether or not one makes the assumption of genuine indeterminism within the quantum domain. Quantum experiments and formalism do not rule out the possibility that electrons, after all, move deterministically along trajectories, being in this respect like their macroscopic cousins, the billiard balls. Yet given that the inability to predict still reigns in the domain of individual quantum systems, faith in indeterminism seems more justified on the basis of the empirical data than faith in determinism. Or is this just a matter of taste?

Bohm's 1952 work - and the work along these lines that followed it - is an important contribution to our knowledge. It is important precisely because, by offering a deterministic alternative as a possibility, it brought out clearly that the commitment to genuine quantum indeterminism is ultimately a matter of faith. But as is well known it did not gain Bohm much respect amongst his colleagues back in 1952 - it was felt by many that he was turning the clock back. It was quite common to mention Bohm as an example of someone who was so attached to determinism that he was unwilling to go along with the indeterministic part of the quantum revolution. Today we can see, in various ways, more appreciation of the 1952 work. There is, for example, a research activity known as "Bohmian mechanics" which develops the approach and finds its deterministic features an advantage rather than a drawback. This development further backs up the impression of Bohm as one of the key supporters of determinism in contemporary physics.

2. Bohm's view of determinism and indeterminism as interwoven categories

The fact about Bohm's attitude towards determinism is, however, more complex than the above discussion suggests. There is no denying that Bohm was eager to prove the possibility of a deterministic description of quantum systems. Yet a study of his more philosophical writings reveals that he was equally keen to understand the role of indeterminism and even genuine indeterminism in nature. This was already brought out in his 1957 book *Causality and Chance in Modern Physics* where he argued that determinism and indeterminism are aspects of a more general law. These ideas were then discussed and developed further in a correspondence he had during 1960-1969 with the American artist Charles Biederman, the first part of which will be published in 1998 by Routledge as the book *Bohm-Biederman Correspondence* (hereafter BB). I believe that these letters provide an important clarification of Bohm's notion of determinism and indeterminism.

The reason why the interaction with Biederman brings out Bohm's views on indeterminism so clearly is the fact that Bohm and Biederman disagreed about the status of indeterminism. For Biederman indeterminism was subjective in the sense that it only arose in the meeting of human beings and nature, while he assumed nature without human beings to be determinate. This prompted Bohm to very clearly spell out in which sense he thought indeterminism to be an objective property of nature. So here we can see Bohm taking the opposite role, that of defending indeterminism rather than questioning it. Indeed, this kind of dialectical approach seems to have been much more important to Bohm than attachment to any particular category such as determinism. Yet even in light of the Bohm-Biederman correspondence it seems that for Bohm the totality of the universe is determined and I guess this defines him as some sort of a determinist. But it is certainly a very weak form of determinism. Consider, for example, the following excerpt by Bohm:

"Of course, you might say that each event is determined necessarily in a complete totality. But this would be trivial. For to give the complete totality, you would have to give each event, so that the determination of that event wouldn't add very much to it. Rather, necessity and determination have real content, only insofar as they operate within fields of abstraction." (BB, p. 164)

Thus, for example, Bohm's 1952 causal interpretation should not be taken as offering in principle a complete description of the totality of the universe at the basic quantum level, but rather as a description of determinism that may operate within a limited "field of abstraction", in this case the quantum domain. Bohm assumed that the quantum determinism he proposed would very likely dissolve into chance in some broader field or sub-quantum domain. The notion of "fields of abstraction" is elaborated earlier in the same letter:

"...the existence of fields of abstraction is characteristic of nature's structure and order. The partial and relative dependence and independence of these fields is also characteristic of nature. Therefore, in each field, there must be contingency and chance. What is chance in one field may be necessary in a broader field and vice versa. But all thought functions in fields. Therefore, we will never get rid of chance and contingency in our thought. But this is not purely the result of our own

way of thinking. For this way of thinking is based on the existence of fields of abstraction in nature... There is genuine contingency, insofar as functioning in fields is valid." (BB, p. 162-3)

The above excerpts have been taken out of a context in which the various concepts have already been discussed from different angles, and thus they are likely to appear somewhat cryptic here. The key point is, however, that determinism and indeterminism are for Bohm complementary concepts that obtain a real, as opposed to trivial, meaning only when we are discussing a part or a partial field of the totality of the universe. Dividing the universe into parts has always the risk of introducing dualisms which seem impossible to interweave back together, thus creating an unfortunate fragmentation to our world view - the dualism between mind and body being perhaps the most notorious case. Yet if we do not introduce divisions, what is there to say? Bohm's dialectical approach is perhaps best summarized in the following excerpt:

"The truth about the cosmos is to be asserted by first asserting its unity, by then asserting its duality in terms of opposing categories, and then showing the interwoven unity of two duals." (BB, p. 50)

In this way determinism and indeterminism are for Bohm an example of opposing categories which are both needed for a non-trivial understanding of the universe. It is in understanding their mutual relationship, how they interweave, that we obtain a non-trivial understanding of natural process. Other categories that he uses in a similar way throughout the correspondence with Biederman include actuality and possibility, finite and infinite, law and lawlessness and regularity and irregularity.

3. Interweaving mind and matter

Perhaps this kind of dialectical method would be a useful addition to many contemporary debates, say in the philosophy of mind. The mind-body problem indeed becomes sharp by Descartes asserting the duality of mind and matter in terms of opposing (mutually exclusive) categories: non-extended, thinking and indivisible mental substance is separated from extended, non-thinking and divisible material substance. The fact that it has been difficult to show the "interwoven unity" of the Cartesian opposites does not mean that this could not work for some other way of distinguishing the mental and the physical. Contemporary philosophy of mind, however, often sees dualism as a bad thing and opts for some form of materialism with the unfortunate result that in most theories of the mind the mind is left out, as Searle (1992) sarcastically points out. Perhaps a different attitude in which one is not trying to eliminate or reduce one category to obtain monism would be more fruitful. Searle's own "biological naturalism" starts from an outright denial of the dualism between mind and matter, insisting that the mind is a biological category. Yet it may be Searle's inability or unwillingness to distinguish the mental from the physical in a satisfactory way at the start of his approach which leaves some of his readers unconvinced. We are given the "right answer" ("mind is a biological category") at the very start of the inquiry but we do not understand how it is a biological category. In Bohmian terms, the Searlian mind doesn't interweave properly with the Searlian brain. Bohm himself discussed the mind-matter issue with pairs of concepts

such as implicate vs. explicate, soma vs. significance and subtle vs. manifest. He also proposed ways of seeing how these categories interweave: the "explicate", for example can be seen as a special case of the "implicate". I believe that a further study of such categories (or similar ones yet to be developed) and their relationship is likely to throw some new light upon even such a perennial problem as the mind-matter one.

Bibliography

Bohm, D. (1952) "A Suggested Interpretation of the Quantum Theory in Terms of Hidden Variables I & II", Phys. Rev., 85, no. 2, 166-193, Republished in Quantum Theory and Measurement, Ed. by J. A. Wheeler and W. H. Zurek, 369-96, Princeton University Press, (1983)

Bohm, D. (1984) Causality and Chance in Modern Physics, London, Routledge and Kegan Paul (new edition of work first published by RKP in 1957).

Bohm, D. (1980) Wholeness and the Implicate Order, London, Routledge and Kegan Paul.

Bohm, D. and Biederman, C. (1998), Bohm-Biederman Correspondence. Volume One: Creativity and Science, Edited by Paavo Pylikkanen, Routledge, London.

Bohm, D. and Hiley B.J. (1993), The Undivided Universe: An Ontological Interpretation of Quantum Mechanics, Routledge, London.

Pylikkanen P. et. al eds. (1997) Brain, Mind and Physics, IOS Press, Amsterdam.

Searle, J. (1992) The Rediscovery of the Mind, MIT Press, Cambridge, Mass.

Developed excerpts from *Mind, Matter and Active Information*

Paavo Pylkkänen

Chapter 1

Background: *Mind, Matter and Active Information: The Relevance of David Bohm's Interpretation of Quantum Theory to Cognitive Science* by P. Pylkkänen is a study that was originally published in 1992 as a monograph, report 2/1992 in the series Reports from the Department of Philosophy, University of Helsinki (ISBN 951-45-6190-2), copyright with the author. I have selected and modified some parts of it (chapters 1 and 2) to be presented at the web course *Quantum approaches to understanding the mind* (University of Arizona 1999). As the text is under development, it contains "first drafts". I have marked additions with the text "(added in 1999)".

CHAPTER 1: INTRODUCTION

This study explores the relevance of David Bohm's interpretation of quantum theory to the philosophy of cognitive science. My first aim is to evaluate Bohm's new mind-matter theory, based upon his ontological interpretation of the quantum theory. In particular, the aim is to explore whether Bohm is justified in extending his notion of active information to apply at the quantum level, and in using this extension as an integral part in his proposed new mind-matter theory (Bohm and Hiley 1987; Bohm 1990).

The second main aim is to explore the implications of Bohm's mind-matter theory to relevant topics in current philosophy of cognitive science (e.g. the problem of mental causation, Crane 1992, Kim 1990). More specifically, the second aim is to answer the following question: *If it in fact were the case that the notion of active information can be justifiably applied in quantum physics (thus integrating Bohm's current mind-matter theory with fundamental laws of physics), what (if any) implications would this have to the philosophy of cognitive science.*

In order to carry out the first aim I will discuss extensively the context in which Bohm extends active information to the quantum level: Bohm's well known ontological interpretation of the quantum theory. It will be argued that this interpretation in some ways provides a more *intelligible* account of the quantum level than most other interpretations (c.f. Cushing 1991). At the same time there remain problems with it, and it is not yet clear how generally the current form of the interpretation applies (although Bohm and Hiley claim (and this has not been refuted to my knowledge) that it is consistent with all known experimental results of quantum theory and relativity). This means that it is premature to arrive at a final judgement about how generally active information applies in the description of nature at the quantum level (although it does apply in all known experimental situations). The verdict of this part of the study is that although it does seem reasonable to extend the notion of active information to the quantum level, we do not know yet how important a place this notion will have in the world view that may one day combine the known physics of today (especially quantum theory and general relativity).

Not being able to arrive at a conclusive answer regarding the general applicability of active information in fundamental physics, my second main aim takes the form of trying to answer the following conditional: *if the notion of active information would apply generally in fundamental physics, what (if any) implications would this have for the philosophy of cognitive science?* The clarification of this question can be important in two ways. On the one hand, if it turns out that there are no interesting implications, then

even if the proposal that active information applies in the quantum level were true, it would not have the kind of significance that it may appear to have. On the other hand, if there are important implications, then we have additional motivation for examining the validity of the proposal.

In order to carry out the second aim I will consider the relevance of Bohm's mind-matter theory to a number of issues in current philosophy of cognitive science, focusing on the mind-body problem and on the current philosophical debate about the problem of *mental causation* and to the assumptions that give rise to this problem (according to Crane 1992):

- 1) anomalism of the mental,
- 2) causal closure of the physical,
- 3) the causal inefficacy of the functional,
- 4) the causal inefficacy of the semantic.

I will also make more general remarks about the relevance of Bohm's theory to other issues in the philosophy of cognitive science, including the following:

- 1) Functionalism and multiple realizability.
- 2) Natural kinds in special sciences vs. physics (Fodor 1975).
- 3) Can the laws of physics and those of the other sciences contradict each other (Field 1989)?
- 4) The physics-involving premises of Fodor's attempts to naturalize intentionality or to solve Brentano's problem (Fodor 1987, 1991).
- 5) Does adherence to a principle of the unity of science require us to eliminate the concept of meaning and other folk psychological concepts (P.M. Churchland 1989)?
- 6) Putnam's thesis about the autonomy of our mental life.

As the main result of this study I suggest that the extension of the notion of active information to the quantum level, if correct, would be relevant to current views in the philosophy of cognitive science. This extension seems to raise a number of interesting possibilities. At least initially it would seem to have the following advantages:

- 1) It means that active information is an important unifying principle in science. Those who value the *unity of science* (not necessarily in a reductionistic sense) might be tempted by this world view (an alternative to eliminative materialism).
- 2) It helps to make *intelligible* the appearance of information (aboutness, meaning) in more developed form in levels higher than the quantum level. It seems that the existence of "intrinsic information" is more intelligible in Bohm's quantum universe than it is in Newton's (and Descartes') world view and in the other cosmologies of modern physics (Bohm's approach may thus provide an alternative to Fodor's etc. way of naturalizing intentionality).
- 3) It implies one suggestion about how mental states (characterized in terms of their information content) can be genuinely *causally efficacious*. According to Bohm's 1990 mind-matter theory, the information content in the subtle level of human mental states reaches, via a series of levels of information, ultimately the quantum field of information of the particles in the brain, and can thus bring about a manifest and visible movement of the

body (via the amplification of these effects; the brain would thus be analogous to a quantum measuring apparatus in this respect). In a reverse process, perception is seen to start from the manifest level of the movement of the particles which affects the quantum field and which in turn affects higher level fields, ultimately reaching the subtle levels of mind. This would mean that the basic theory of current physics *implies* a world view in which mental states are causally efficacious. This turns the current debate about mental causation upside down, for it is often precisely by appealing to authority of "physics" or "physicalism" that one is led to deny the causal efficacy of the mental *qua* mental (an alternative to epiphenomenalism).

At the same time there are, at present, rather serious problems with Bohm's mind-matter hypothesis involving quantum theory:

- 1) It is not clear whether particles in the brain could be sensitive to quantum effects (because of the conditions, e.g. temperature).
- 2) The ontological interpretation lacks an account of how the particles could affect the field. (Added in 1999: Jack Sarfatti has worked on this part)
- 3) Apart from Fröhlich's (e.g. 1986) work, there are as yet few concrete proposals about how quantum type of organization could manifest itself in biological organisms (thus justifying the postulation of intermediate levels of fields of active information between the quantum and mental levels; c.f. however Sheldrake 1988). (Added in 1999: As the web course shows, there is a wealth of approaches currently on offer)

Whether or not these problems can be solved cannot be settled within this study; they are partly empirical questions and it may take a long time before they are answered. Our concern is more to evaluate whether this set of problems is relevant to cognitive science. To do this we have to discuss the implications that the hypothesis of extending active information to the quantum level, if correct, would have for cognitive science.

Let me say a few words about how the aims of this study have come about. The primary question underlying my research over the past few years has been the possible relevance of Bohm's general mind-matter theory and his theory of meaning to current views in the philosophy of cognitive science. The reason why this study gives such a central emphasis to understanding the extension of the notion of active information to the quantum level (and thus to understanding the ontological interpretation of the quantum theory, which is crucial to this extension) is that the philosophy of cognitive science tends to be physicalistic and places a high value upon the laws and ontology of physics. Bohm's approach implies that the fundamental level of physics is *qualitatively different* from what it is currently assumed to be. This may imply the need to reconsider one's physicalism.

In order to understand the extension of active information to the quantum level, we have to understand the ontological interpretation of the quantum theory. This in itself is a vast topic which could be approached in many ways. I have considered issues that are either directly relevant to understanding the role of active information at the quantum level, or else relevant to the credibility of the ontological interpretation more generally (which latter issues are crucial for the applicability of active information in the quantum level,

for if the ontological interpretation is not plausible, neither is the notion of quantum theoretical active information. I have thus not restricted myself to constantly discussing only those issues which are directly relevant to the notion of active information. In fact, Chapters 4 and 5 can be seen as exploring the nature and validity of the ontological interpretation more generally, whereas Chapter 6 analyzes the notion of active information from different points of view. Chapter 7 then sketches generally the implications of Bohm's mind-matter theory to various issues in the philosophy of cognitive science, while Chapter 8 takes a closer look at implications for the problem of mental causation, and in particular for the problem of the causal inefficacy of the functional.

It may seem strange that I have focused rather little on the analysis of the concept of active information itself as well as on its relation to other notions of information. I have in a previous study discussed active information in relation to other current notions of information (MSc thesis, University of Sussex, 1985). The historical background of Bohm's concept of active information as well as its relation to other concepts of information have also been more thoroughly discussed by Gordon Miller in his dissertation, which also examined the relevance of active information in perceptual psychology (PhD, Rutgers University, 1987). My main contribution here in this regard will be to evaluate the plausibility of quantum theoretical active information by discussing the ontological interpretation, as well as to study those aspects of active information which are most important from the point of view of the mind-body problem.

Developed excerpts from *Mind, Matter and Active Information*

Paavo Pyykkänen

Chapter 2

Background: *Mind, Matter and Active Information: The Relevance of David Bohm's Interpretation of Quantum Theory to Cognitive Science* by P. Pyykkänen is a study that was originally published in 1992 as a monograph, report 2/1992 in the series Reports from the Department of Philosophy, University of Helsinki (ISBN 951-45-6190-2), copyright with the author. I have selected and modified some parts of it (chapters 1 and 2) to be presented at the web course *Quantum approaches to understanding the mind* (University of Arizona 1999). As the text is under development, it contains "first drafts". I have marked additions with the text "(added in 1999)".

CHAPTER 2: QUANTUM THEORY AND THE MIND-BODY PROBLEM: A SUMMARY OF BOHM'S VIEWS

2.1 Introduction

In this study we are concerned with two questions: firstly, in evaluating Bohm's mind-matter theory and, in particular, in understanding the validity of the ontological interpretation and the use of active information in its context. Secondly, we are interested in the possible relevance of this theory to the philosophy of cognitive science and the problem of mental causation in particular. In this chapter my aim is to introduce Bohm's mind-matter theory and to discuss its relation more generally to other similar views. We will then in the following chapters examine possible objections to his scheme. Some might claim that there are *prima facie* reasons for thinking that quantum theory is not relevant to the mind-body problem, reasons which have to be taken into account by anyone proposing otherwise, including Bohm. We will thus examine some such *prima facie* reasons and examine whether Bohm's approach is touched by them (Chapter 3). We will also discuss the nature and plausibility of the ontological interpretation itself; its advantages and disadvantages and some criticisms levelled against it (Chapters 4 and 5). For if this interpretation should turn out to be incorrect, it is not clear whether active information can be extended to the quantum level, and thus the current form of Bohm's mind-matter theory cannot be sustained.

In discussing the ontological interpretation we will give specific attention to the question of whether it is justified and reasonable to make the radical hypothesis that active information applies at the quantum level. What are the reasons for postulating this? What does this concept explain? Could we do without it just as well? Is it just a matter of taste, or do the experiments imply the need for its postulation? We will then try to understand the nature of active information more deeply. Does it imply a new concept of causality? In what sense is it information? Does it have representational or intentional properties? But we are running ahead of ourselves. Let us first have a look at Bohm's views on the relation of mind and matter, the ontological interpretation of quantum theory and the notion of active information.

2.2 Bohm's 1990 paper on mind and matter

(Added in 1999 for the web course): this chapter discusses Bohm's 1990 paper in detail and also makes many quotations of it. It is, however, recommended that you read the original paper in the web address first & can in that way compare your own reading of it with the remarks below. (end of 1999 insert)

Bohm has discussed the problem of the relation of mind and matter throughout his recent work and especially in his 1980, 1985, 1986, 1987 (with David Peat), 1989 and 1990. I will here concentrate on his latest formulation on this topic, the 1990 paper "A New Theory of the Relation of Mind and Matter" which was published in the journal *Philosophical Psychology*. Let us first consider what he says in the abstract:

The relationship of mind and matter is approached in a new way in this article. This approach is based on the causal interpretation of the quantum theory, in which an electron, for example, is regarded as an inseparable union of a particle and a field. This field has, however, some new properties that can be seen to be the main sources of the differences between the quantum theory and the classical (Newtonian) theory. These new properties suggest that the field may be regarded as containing objective and active information, and that the activity of this information is similar in certain key ways to the activity of information in our ordinary subjective experience. The analogy between mind and matter is thus fairly close. This analogy leads to the proposal of the general outlines of a new theory of mind, matter, and their relationship, in which the basic notion is participation rather than interaction. Although the theory can be developed mathematically in more detail, the main emphasis here is to show qualitatively how it provides a way of thinking that does not divide mind from matter, and thus leads to a more coherent understanding of such questions than is possible in the common dualistic and reductionistic approaches. These ideas may

be relevant to connectionist theories and might perhaps suggest new directions for their development. (1990: 271)

This summary brings succinctly together the issues with which we will be concerned throughout this study. Bohm emphasizes that his new approach is *based on* his causal (or as he later preferred to call it, ontological) interpretation of the quantum theory. We are told the basics of Bohm's quantum ontology: each individual system is simultaneously both a particle and a field (as opposed to the wave-particle duality of the "conventional" interpretation of quantum theory according to which an individual quantum system is to be described either as a wave or a particle at any one moment (but not as both simultaneously), depending on the experimental situation). We are also told that the new properties of the quantum field ? are the essential difference between the quantum theory and Newtonian theory. This point will be crucial as we try to understand the relation between the classical level and quantum level - a question which has been very difficult to understand in the conventional interpretations of the quantum theory.

The next statement in the abstract expresses a point central to this study, as Bohm goes on to say that the new properties of the quantum field suggest that the "*...field may be regarded as containing objective and active information, and that the activity of this information is similar in certain key ways to the activity of information in our ordinary subjective experience*". This statement expresses the essence of Bohm's 1990 mind-matter theory and much of what we will do in this study will be to unpack its implications. We want to understand

- 1) the nature of this field and its status in modern physics, thus the nature and plausibility of the ontological interpretation;
- 2) what it is in the properties of the quantum field that prompts Bohm to regard the field as containing objective and active information;
- 3) what Bohm means by information and by saying that it is *active* and *objective*;
- 4) whether Bohm is justified and reasonable in saying that the activity of information at the quantum level is in certain *key* ways similar to the activity of information in our ordinary subjective experience;
- 5) even if such similarity were to be found, what would this imply.

Having said something about these questions, we will then explore the possible relevance of Bohm's mind-matter theory to issues in cognitive science, and to the debate about mental causation in particular. It will become clear that to emphasize the relevance to mental causation is not arbitrary, for Bohm's theory may (if correct) indeed have something interesting to say about this. Another topic in current philosophy of cognitive science to which Bohm's theory seems to be relevant is the question about the place of information in nature and the use of the concept of information to provide an understanding of intentionality or the "place of meaning in the world order" (Fodor 1987, 1990; Dretske 1981).

In his abstract, Bohm further points out that there is a fairly close analogy between mind and matter. This means that the way the information in the quantum field acts to organize the movement of the particle is fairly closely analogous to the way information in our subjective experience acts to guide less subtle, material levels and ultimately moves the body. Guided by this analogy Bohm then proposes the general outlines of a new theory of mind, matter and their relationship in which the basic notion is *participation* rather than interaction. This theory is put forward as "... a new way of thinking, consistent with modern physics, which does not divide mind from matter, the observer from the observed, the subject from the object" (1990: 271).

This emphasis on participation has to do with quantum wholeness, which also plays a key role in measurement at the quantum level. In quantum measurement the observing apparatus and the observed system can be said to participate in each other. This means that they are indivisibly connected during the measurement and so it is not possible to ascribe the result of the measurement to the observed system alone. The observing apparatus participates in bringing about the state of the observed system which is being "measured".

Bohm emphasizes that his new mind-matter theory is *consistent* with modern physics. Hautamäki (private communication) has pointed out that such consistency may not be all that difficult to establish, if it is understood that in order to obtain consistency all that is required is to avoid contradiction. Thus even theories which have very little in common can be consistent with each other as long as there are no contradictions between them. It would make Bohm's new way of thinking more interesting if one, for example, could say that it is *implied* by modern physics. Bohm says that his new mind-matter theory is "based on the causal interpretation" and this is close to saying that the former is implied by the latter.

2.3 Has the Cartesian view of matter as a substance been given up by modern philosophers?

In the introductory part of his 1990 paper Bohm expresses his motivation for developing a new mind-matter theory. He recognizes that Cartesian substance dualism is not workable without the intervention of God and that the latter idea is no longer considered as a valid basis for a philosophical argument. He also distances himself from reductive materialism or idealism:

This article aims at the development of a different approach to this question, which permits of an intelligible relationship between mind and matter without reducing one to nothing but a function or aspect of the other (such reduction commonly takes the forms of materialism which reduces mind, for example, to an "epiphenomenon" having no real effect on matter, and of idealism, which reduces matter to some kind of thought, for example, in the mind of God). (1990: 272)

From the point of view of the current debate on mental causation it is interesting to note that Bohm is not satisfied in seeing mind as an epiphenomenon, having no real effect on matter. Thus his theory can be seen as an attempt to develop a non-epiphenomenalist view of mind, according to which mind has a real effect upon matter. In Chapter 8 we will examine whether Bohm can be considered as having succeeded in this - or whether his approach can be considered to be on a promising track. It is also important to note that Bohm rejects reductive idealism, for his view may *prima facie* seem close to Hegel's objective idealism (see 4.5).

In relation to the history and the current state of the mind-body problem, perhaps the most significant aspect of Bohm's theory is that he challenges the Cartesian way of understanding inanimate matter:

... the quantum theory, which is now basic, implies that the particles of physics have certain primitive mind-like qualities which are not possible in terms of Newtonian concepts (though, of course, they do not have consciousness). This means that on the basis of modern physics even *inanimate* matter cannot be fully understood in terms of Descartes' notion that it is nothing but a substance occupying space and constituted of separate objects. (1990: 272)

Bohm is surely not the only one to criticize the Cartesian notion of matter as extended substance on the basis of modern physics. Consider, for example, Paul Churchland:

... the basic principle of division used by Descartes is no longer as plausible as it was in his day. It is now neither useful nor accurate to characterize ordinary matter as that-which-has-extension-in-space. Electrons, for example, are bits of matter, but our best current theories describe the electron as a point-particle with no extension whatever (it even lacks a determinate spatial position). And according to Einstein's theory of gravity, an entire star can achieve this same status, if it undergoes a complete gravitational collapse. If there truly is a division between mind and body, it appears that Descartes did not put his finger on the dividing line. (1984: 9)

At the same time the following extract from Jaegwon Kim shows that he considers that philosophers have given up only the notion of mental substance:

... mental substance having entirely vanished from the scene, it is now difficult for us even to formulate the Cartesian problem. In fact, the rejection of mental substance was the simplest solution to that problem, although I don't know whether historically this played a role in the demise of the Cartesian soul. (Kim 1990: 36)

Consider also the following quotation from Arthur Danto's recent book:

Philosophy ... suffers from very early decisions as to what the structure of the world must be if certain facts are to be understood. Even when one would no longer frame theories as once was done, the problems with which philosophers deal rest on archaic divisions they repudiate. How much of the so-called mind-body problem would remain were we to erase, really erase, certain concepts that gave it form but which have long fallen into disuse, is difficult to say. One of the concepts, for example, is that of substance. (1989: 210)

Regarding Danto's above statement, perhaps the way Bohm questions Descartes' notion of inanimate matter may really help to "erase" the Cartesian habit of thinking of matter as an independent substance. For, one might ask, what else is the almost universal adherence to vaguely defined "physicalism" in the current philosophy of cognitive science than a tacit commitment to the old, Cartesian mechanical concept of matter as a substance? Indeed, one might suggest that the current debate about mental causation and the *closeness* to epiphenomenalism with regard to the mental of practically every view in cognitive science is just a sign that almost all participants in the debate still hold onto something like the Cartesian notion of the physical. Kim points out that all parties to the present debate about mental causation adhere to basic physicalistic

assumptions (1990: 45). The question is whether these physicalistic assumptions essentially amount to a Cartesian view of matter as a substance, perhaps modified with ideas from Newtonian physics.

Thus, although Bohm does not relate his mind-matter theory to the current debates in the philosophy of cognitive science in any detail, in this study we will show that in setting out to challenge the Cartesian view of inanimate matter and especially in trying to formulate an alternative view of matter implied by modern physics, he is indeed able to contribute in an important way to understanding the nature of mind, matter and their relation.

2.4 Is the implicate order common to both matter and mind?

In the second section of his 1990 paper Bohm gives a brief summary of his previous ideas on the mind-matter relation based on his earlier work in physics, as they were presented in his book *Wholeness and the Implicate Order* (1980). The key notion here is that of the enfolded or implicate order, which he originally developed when he was trying to understand relativity and quantum theory on a basis common to both. The essential feature of this notion was that

... the whole universe is in some way enfolded in everything and that each thing is enfolded in the whole. From this it follows that in some way, and to some degree everything enfolds or implicates everything, but in such a manner that under typical conditions of ordinary experience, there is a great deal of *relative* independence of things. The basic proposal is then that this enfoldment relationship is not merely passive or superficial. Rather, it is active and essential to what each thing is. It follows that each thing is internally related to the whole, and therefore, to everything else. The external relationships are then displayed in the unfolded or explicate order in which each thing is seen ... as relatively separate and extended, and related only externally to other things. The explicate order, which dominates ordinary experience as well as classical (Newtonian) physics, thus appears to stand by itself. But actually, it cannot be understood properly apart from its ground in the primary reality of the implicate order. (1990: 273)

Thus, already the notion of implicate order challenged the Cartesian view that matter could be understood as extended substance, and proposed an alternative to it. In fact, Bohm's major motivation in developing the implicate order was to go beyond what he calls the Cartesian order. Cartesian order, which is epitomized in the use of Cartesian co-ordinates in physics, is the one feature that physicists have not given up regardless of the revolutions of physics.

The fact that it is possible to create and annihilate elementary particles (as described by quantum field theory) has led many, including Bohm to emphasize that reality is a process. Popper describes this as follows:

Matter is not "substance", since it is not conserved: it can be destroyed, and it can be created. Even the most stable particles, the nucleons, can be destroyed by collision with their anti-particles, when their energy is transformed into light. Matter turns out to be highly packed energy, transformable into other forms of energy; and therefore something of the nature of a *process*, since it can be converted into other processes such as light and, of course, motion and heat. Thus one may say that the results of modern physics suggest that we should give up *the idea of a substance or essence*. They suggest that there is no self-identical entity persisting during all changes in time (even though bits of matter do so under "ordinary" circumstances); that there is no essence which is the persisting carrier or possessor of the properties or qualities of a thing. *The universe now appears to be not a collection of things, but an interacting set of events or processes* (as stressed especially by A.N. Whitehead). (1977: 7; last emphasis added)

Bohm's notion of implicate order also includes an attempt to characterize not only the mutual interdependence of all things with each other and the whole, but also to characterize the universe and its relatively independent parts in terms of the concept of process:

Because the implicate order is not static but basically dynamic in nature, in a constant process of change and development, I called its most general form the *holomovement*. All things found in the unfolded, explicate order emerge from the holomovement in which they are enfolded as potentialities and ultimately they fall back into it. They endure only for some time, and while they last, their existence is sustained in a constant process of unfoldment and re-enfoldment, which gives rise to their relatively stable and independent forms in the explicate order. (1990: 273)

It is clear that the implicate order is particularly suited to describe the creation and annihilation of particles. It takes this as the paradigm of natural philosophy and contains the old Cartesian model as an approximate way of describing things when their existence is sustained.

A basic difficulty with understanding the relation of mind and matter has traditionally been how mind which is non-spatial can have anything to do with matter which is spatial. Modern physics reveals that the notion of matter as spatial and extended is at best an approximation. According to Bohm, what is essential to matter is unfoldment and enfoldment. There is a sense in which matter, too, is non-spatial, or that it cannot be fully understood in spatial terms. If the ground of matter is in something non-spatial (beyond 3 + 1 dimensional space-time) then perhaps the apparent spatiality of matter and the apparent non-spatiality of mind can no longer be considered as *the* essential source of the mind-body problem (c.f. Rorty (1979: 17-22); Feigl (1967: 39-42)). Perhaps the notion of implicate order in terms of which matter can be understood also provides a framework in which to discuss mind. This is exactly what Bohm proposed in the last chapter of his book *Wholeness and the Implicate Order* (1980). Indeed, he says that the implicate order ... applies even more directly and obviously to mind, with its constant flow of evanescent thoughts, feelings, desires, and impulses, which flow into and out of each other, and which, in a certain sense, enfold each other (as, for example, we may say that one thought is *implicit* in another, noting that this word literally means 'enfolded'). Or to put it differently, the general implicate process of ordering is common both to mind and to matter. This means that ultimately mind and matter are at least closely analogous and not nearly so different as they appear on superficial examination. Therefore, it seems reasonable to go further and suggest that the implicate order may serve as a means of expressing consistently the actual relationship between mind and matter, without introducing something like the Cartesian duality between them. (1990: 273)

Thus, in Bohm's view, neither mind nor matter are substances; they are relatively stable and independent aspects of the holomovement; and the concept of implicate order expresses the way they depend upon the holomovement and how they enfold each other. Because mind and matter are not conceived of as substances, the Cartesian mind-matter problem does not arise. Insofar as it is possible to discuss the relationship of different aspects of relatively independent levels in a single underlying process, it may well be possible to understand the relationship of the mental and the physical sides of reality in the framework of the implicate order. Bohm, however, is not satisfied with the implicate order in this regard:

At this stage ... the implicate order is still largely a general framework of thought within which we may reasonably hope to develop a more detailed content that would make possible progress toward removing the gulf between mind and matter. Thus, even on the physical side, it lacks a well-defined set of general principles that would determine how the potentialities enfolded in the implicate order are actualized as relatively stable and independent forms in the explicate order. The absence of a similar set of principles is, of course, also evident on the mental side. But yet more important, what is missing is a clear understanding of *just how* mental and material sides are to be related. (1980: 273 - 274)

Bohm thus outlines three problems: 1) As a theory of the physical the implicate order cannot account for how extended matter manifests from the implicate ground; 2) There is no account of how the potentialities in the mental level result in actualization (an example of this could be making a decision leading to bodily action); 3) The implicate order does not provide a specific account of how mental and material sides are related.

Bohm notes that one has to develop the notion of the implicate order if one wants to have a satisfactory account of the relation of mind and matter. He then says that the ontological (or causal) interpretation of the quantum theory goes a long way toward fulfilling this requirement. And as we have already emphasized, it is in the context of the ontological interpretation that he introduces the notion of active information at the quantum level, and it is this notion that he thinks may provide a clearer understanding of just how the mental and material are related.

It is interesting to notice that not only does Bohm think that the ontological interpretation may help to provide a more specific mind-matter theory, he also thinks that it may help with the difficulties the implicate order has as a theory of the physical (Bohm 1987c). As Hiley and Peat (1987: 18) point out, Bohm has found that certain features of the ontological interpretation (which come out in quantum field theory)

"... offer a guide as to the limitations that must be imposed on the implicate order if it is to produce results that are not contrary to our experience of quantum phenomena".

The fact that the ontological interpretation today appears to provide an important insight to both the mind-body problem and the problem of developing the implicate order as a theory of the physical is rather

curious. For the ontological interpretation has traditionally been considered as representing Bohm's "old ideas" - old both in the sense that he proposed them almost fifty years ago and also in the sense "old-fashioned" in that they introduced the particle as a well-defined entity to the ontology of quantum theory (though presumably this was not offered as a concept upon which one could build a new notion of matter as a substance). The ontological interpretation was by and large rejected by the physics community when it first appeared in 1952 (Pinch 1976, 1977a, 1977b; Bohm and Peat (1987: 87-103); Cushing 1991b; Bell 1987); presumably because it was a disturbing counterexample to many of the claims that were held true by the supporters of the conventional interpretation. Because it introduced the possibility of continuous space-time description of particles and determinism at the quantum level it was also seen as a step back towards Newtonian physics. As time has passed, the ontological interpretation has been reevaluated - not least because John Bell repeatedly insisted that its dismissal in the 1950s was by and large unfounded (see e.g. his 1987; see also Holland 1993, Cushing 1991b and Pinch 1976, 1977a).

It is not yet clear how satisfactorily the ontological interpretation applies to relativistic quantum field theory - it is especially hard to see how it could be made compatible with general relativity. This applies, of course to quantum theory and general relativity in general, and even the union with quantum theory with special relativity is not considered to be satisfactory (Bohm and Hiley (1987: 374); Penrose (1989: 289)). Thus the fate of the ontological interpretation depends at least partly upon the fate of the relation of quantum theory and relativity in general.

At the same time it seems that the ontological interpretation does contain features which may be useful in understanding a number of important issues, such as how to make the implicate order workable in physics and how to provide a theory of the relation of mind and matter. These are no minor advantages, and thus regardless of its limits (which presumably every scientific idea has) the ontological interpretation may turn out to have an important, if not fundamental, role in our current understanding of nature.

In this section we have explored the idea that it might be easier to understand the relation of mind to the matter of modern physics than it is to the matter of Descartes or Newton. We noted that although the implicate order provides a general idea of how to understand the relation of mind and matter, something more is needed. Bohm proposes that the ontological interpretation of the quantum theory goes a long way toward providing a more specific account of just how mind and matter are related, a basis for a non-dualistic mind-matter theory. But before going on to discuss this interpretation, we will say something more general about the quantum theory and its problems in order to understand why it is necessary to develop such an ontological interpretation.

2.5 Quantum theory and its mysterious, confusing features

Quantum theory, perhaps more than any other theory in the history of science has presented philosophical problems of *interpretation* after its completion. Regardless of the fact that the theory was developed as a successfully applicable mathematical formalism already in 1927 and that there has been a vast amount of literature on the foundations of the theory, there still persists a disagreement and perhaps a confusion about the *meaning* of the theory among the researchers specializing in the interpretation of the theory, as well as among other physicists, philosophers of science and the wider academic community.

Because quantum theory is presently the most fundamental theory of matter and because modern science and philosophy is by and large physicalistic, this confusion about interpretation amounts to a confusion in the modern scientific *world view*. In other words, the scientific community has no clear and commonly accepted *ontological* notion of the physical world. For example, philosophers of mind and cognitive science widely announce themselves to be "physicalists", implying that they believe in the existence of whatever the best theories of physics say. At the same time physics is in a crisis and does not really have a clear idea of the nature of existence at the physical level.

Bohm and Hiley outline four basic points that are not clear about the meaning of the quantum theory:

1. [Measurement problem] Though the quantum theory treats statistical ensembles in a satisfactory way, we are unable to describe individual quantum processes without bringing in unsatisfactory assumptions, such as the collapse of the wave function.
2. [Non-locality] There is by now the well known non-locality that has been brought out by Bell in connection with the EPR experiment.
3. [Wave-particle duality] There is the mysterious 'wave-particle duality' in the properties of matter that is demonstrated in a quantum interference experiment.
4. [Nature of a quantum system] Above all, there is the inability to give a clear notion of what the reality of a quantum system could be. (1992: Chapter 1: 1)

One should note that although these points are not well understood, they are at the same time considered to be the radical, non-Newtonian new features of the quantum theory, features behind such concepts as "quantum wholeness" which were emphasized e.g. in Niels Bohr's thinking.

Especially regarding the impossibility to give an account of individual quantum systems, Bohr's interpretation assumes that this is somehow an inherent aspect of the quantum theory. Bohr is well known for his view that an observation involving quantum effects has to be understood as a single *phenomenon*, which is a whole that is not further analyzable:

For Bohr this implies that the mathematics of the quantum theory is not capable of providing an unambiguous (i.e., precisely definable) description of an individual quantum process, but rather, that it is only an algorithm yielding statistical predictions concerning the possible results of an ensemble of experiments. Bohr further supposes that no new concepts are possible that could unambiguously describe the reality of the individual quantum process. Therefore, there is no way intuitively or otherwise to understand what is happening in such processes. Only in the Newtonian limit can we obtain an approximate picture of what is happening, and this will have to be in terms of the concepts of Newtonian physics. (Bohm 1990: 275)

The starting point of the ontological interpretation is thus to question Bohr's assumption that no conception of the individual quantum process is possible. Bohm simply assumes that

... the electron, for example is a particle which follows a well defined trajectory (like a planet around the sun). But it is always accompanied by a new kind of quantum field. (1990: 276)

Having provided such a notion it then turns out that one may be able to provide an intelligible resolution of the other three above mentioned "paradoxes" of quantum theory, i.e. wave-particle duality, non-locality and the measurement problem, without making any extra assumptions. In the ontological interpretation these three obscure aspects become *intelligible* in the sense that they are a natural consequence of the properties of individual quantum systems, and these properties, although they are new and strange, we can nevertheless understand. It would thus appear that by "resolving" one of the above four difficulties by simply making an assumption or hypothesis, Bohm is then able to resolve the other three difficulties. This can be contrasted with a situation in which one assumes something new for each problem that has to be explained. Bohm has emphasized that *every* theory has to make *some* assumptions:

... the whole point of science is to begin with some assumptions and see if you can explain a wide range of things from a few assumptions. This enables you to *understand* in the sense that far more things are explained than you have assumed. (1989: 69)

I thus propose that we look at Bohm's ontological interpretation as an example of following such scientific method. Bohm and Hiley begin their book *Undivided Universe* (1993) by giving a preliminary list of things that are unclear ("obscure and confused") about quantum theory (implying that these need clarification or have to be explained if possible). The strategy of the ontological interpretation is then to resolve one unclarity (i.e. that concerning the reality of a quantum system) by an assumption or hypothesis, and explain the others from this assumption. It is in this sense that the ontological interpretation, although it contains strange features, nevertheless can be seen as an attempt to resolve paradoxes and confusions that exist in the conventional interpretation. This to me seems to be the strategy of the ontological interpretation; it is, of course another matter how successful the interpretation can be considered in realizing this strategy.

2.6 Bohm's ontological interpretation of the quantum theory: explaining wave-particle duality

Let us now consider Bohm's hypothesis about the nature of the quantum system and see in which sense it makes intelligible wave-particle duality and non-locality (we will discuss quantum measurement in Chapter 5, section 3.4). In the conventional interpretation, the fundamental laws of the quantum theory are expressed with the aid of a *wave function* ψ . One of the crucial steps in developing the conventional interpretation was Max Born's suggestion that the wave function has to be given a probability interpretation. It is thought of as a "probability wave" instead of being seen as a real physical wave. It tells us, for example, the probability of *finding* a particle at a certain point in a measurement (as opposed to giving us a probability of there *being* a particle at that point regardless of measurement). Bohm's interpretation challenged profoundly this standard interpretation of the wave function:

The basic assumption was that the electron *is* a particle, acted on not only by the classical potential, V , but also by the quantum potential, Q . This latter is determined by a new kind of wave ψ that satisfies Schrödinger's equation. This wave was assumed, like the particle, to be an *independent actuality that existed on its own, rather than being merely a function from which the statistical properties of phenomena could be derived*. However, I showed on the basis of further

physically reasonable assumptions that the intensity of this wave is proportional to the probability that a particle *actually is* in the corresponding region of space (and is not merely the probability of our observing the phenomena involved in *finding* a particle there). So the wave function had a double interpretation - first as a function from which the quantum potential could be derived and secondly, as a function from which probabilities could be derived. (1987c: 36, long emphasis added.)

It is thus the quantum wave field ?, ("wave function", "quantum wave", or "quantum field" being alternative names) that gives rise to the quantum potential acting on the particle. Although this field is assumed to be real (ontologically, not mathematically!), it differs in interesting respects from other known fields. First of all, the strength of the quantum potential does not depend on the intensity of the wave associated with the electron, it depends only upon the *form* of the wave. A key implication of this is that the effect of the quantum potential can be large even when the wave has spread out by propagation across large distances and its intensity has thus become very weak. In other words,

... even a very weak quantum field can strongly affect the particle. It is as if we had a water wave that could cause a cork to bob up with full energy, even far from the source of the wave. Such a notion is clearly fundamentally different from the older Newtonian ideas. For it implies that even distant features of the environment can strongly affect the particle. (1990: 276)

This feature is illustrated when we consider how Bohm's model accounts for the well known two-slit experiment. Before doing that let us first state the problem in a standard way, following Feynman (1965) and Pagels (1982).

(will be added shortly)

Fig. 1. The behaviour of bullets, water waves and electrons in the two-slit experiment. It is easy to understand why bullets and water waves arrive at the places they do in the detector placed behind the slits. It is, however, very strange that the electrons should form an interference distribution (E) when two slits are open. (The figure is from Pagels 1982)

The two upper pictures remind us how particles and waves behave according to classical physics when they go through a partition which contains two adjacent slits, one or both of which may be open. The first picture illustrates, using bullets as an example, how particles behave in the situation. P1 describes the distribution of bullets in a wall when only slit 1 is open, P2 describes it when only slit 2 is open and P describes the distribution when both slits are open. It is obvious that the opening of the second slit increases the possible places at which bullets may arrive. P is simply the sum of P1 and P2.

In the second picture water waves illustrate how classical waves behave in a corresponding situation. W1 describes the intensity distribution of waves on a detector when just one slit is open. This resembles P1, which we got with bullets in the corresponding situation. The situation is the same with W2 and P2. If we now open both slits, we get an intensity distribution W which is different from P, for W is not a sum of W1 and W2. According to W strong waves arrive at certain regions whereas in between these no waves at all arrive. The reason for this is that in some places the waves coming from slits 1 and 2 strengthen each other (they interfere constructively when they are in phase), whereas in other places they cancel each other out (they interfere destructively when they are out of phase). In the two-slit experiment we get interference fringes with classical waves.

What happens with quantum systems? In the lowest picture electrons "boil" out of a hot filament, and they move through the slits toward a photographic plate. The picture reveals that the electrons have astonishing wave-like properties. The intensity distribution E of an ensemble of electrons is similar to the water wave distribution W in the corresponding situation. In this sense the electrons have wave-like properties. On the other hand, individual electrons register to the photographic plate at a single spot, as if they were localized particles.

If we now close slit 2, electrons arrive at the photographic plate according to E1. The key point here is that electrons now can arrive at regions at which they never can arrive when both slits are open. Thus, the opening of the second slit prevents the electrons from arriving at regions where they can arrive with just one slit open. But how does the electron "know" whether the second slit is open or not? It is clear that one cannot explain the situation by saying that the electron has only a particle aspect.

Paradoxes like these make it *prima facie* reasonable that in the usual interpretation of the quantum theory the electron is described by a wave function. Bohm gives the following account of the way the usual interpretation describes the situation in terms of a wave model:

... when the wave passes through the slit system, it is modified by interference and diffraction effects, so that it will develop a characteristic intensity pattern by the time it reaches the position measuring instrument. The probability that the electron will be detected between x and $x + dx$ is $|\psi(x)|^2 dx$. If the experiment is repeated many times under equivalent initial conditions, one eventually obtains a pattern of hits on the photographic plate that is very reminiscent of the interference pattern of optics. (Bohm 1952: 173)

He then goes on to give reasons for why he thinks that this wave-particle duality is left unclear in the conventional interpretation (I am here using Bohm's famous 1952 papers, both in order to bring out how the ontological interpretation has developed, and in order to show that many relevant ideas can be found already in these early papers):

In the usual interpretation of the quantum theory, the origin of this interference pattern is very difficult to understand. For there may be certain points where the wave function is zero when both slits are open, but not zero when only one slit is open. How can the opening of a second slit prevent the electron from reaching certain points that it could reach if this slit is closed? If the electron acted completely like a classical particle, this phenomenon could not be explained at all. Clearly, then, the wave aspects of the electron must have something to do with the production of the interference pattern. Yet, the electron cannot be identical with its associated wave, because the latter spreads out over a wide region. On the other hand, when the electron's position is measured, it always appears at the detector as if it were a localized particle. (1952: 173)

The electron cannot be thought of as only having a particle nature; neither can it be understood as only having a wave nature. The question then arises whether it is possible to provide a *model* of the electron which could provide an intelligible account of what happens in, say, the two-slit experiment. Is such a model provided by the conventional interpretation? Bohm tries to clarify this question as he continues his 1952 discussion of the two-slit experiment:

The usual interpretation of the quantum theory not only makes no attempt to provide a single precisely defined conceptual model for the production of the phenomena described above, but it asserts that no such model is even conceivable. Instead of a single precisely defined conceptual model, it provides ... a pair of complementary models, *viz.*, particle and wave, each of which can be made more precise only under conditions which necessitate a reciprocal decrease in the degree of precision of the other. (1952: 173)

According to the usual interpretation it thus seems that we *can* say something about what happens in, say, the two-slit experiment. We can use the wave model (but not the particle model) as long as we talk about the electron as moving through the slit system; whereas if we want to talk about its registration in the photographic plate, we have to use the particle model (and drop the wave model). It is as if the electron is a wave when it moves and then suddenly becomes a particle when it interacts with matter:

Thus, while the electron goes through the slit system, its position is said to be *inherently ambiguous*, so that if we wish to obtain an interference pattern, it is meaningless to ask through which slit an individual electron actually passed. Within the domain of space within which the position of the electron has no meaning we can use the wave model and thus describe the subsequent production of interference. If, however, we tried to define the position of the electron as it passed the slit system more accurately by means of a measurement, the resulting disturbance of its motion produced by the measuring apparatus would destroy the interference pattern. Thus, conditions would be created in which the particle model becomes more precisely defined at the expense of a corresponding decrease in the degree of definition of the wave model. When the position of the electron is measured at the photographic plate, a similar sharpening of the degree of definition of the particle model occurs at the expense of that of the wave model. (Bohm (1952: 173-174); emphasis added.)

This, presumably is the "official" version of the two-slit experiment, and it was widely thought (and perhaps still is) that no other explanation is possible (insofar as the above can be considered as an *explanation* in the first place, for explanations, one may think, ought to make the phenomenon to be explained *intelligible*, and the question is whether complementarity is intelligible).

Having gone through the "explanation" of the two-slit experiment in terms of the conventional interpretation and complementarity, one may feel that it amounts more to a statement that an intelligible explanation for the situation cannot be given. The nature of the individual quantum system is left as a mystery. It is not a wave, it is not a particle, it is not their combination; what is it then? The "answer" to

this question has been to say that the question is meaningless. But how do we know that it is meaningless? Presumably only because we have made some assumptions (those upon which the conventional interpretation is based) that render it meaningless, and taken those assumptions to be true. But assumptions can always be questioned, and presumably never proved to be absolutely true. This was the line of Bohm's approach. If one finds oneself in an unintelligible situation, it is reasonable to examine one's assumptions about the situation. Bohm questioned the assumption that it is not possible to provide a single precisely definable conceptual model in terms of which to describe quantum phenomena, including the two-slit experiment. He did this by providing a counterexample, i.e. by proposing such a model. Let us now see how Bohm's model according to which an electron is simultaneously a particle *and* a new type of field (the latter being described by the wave function ψ) gives an account of the two-slit experiment. Consider an electron coming from the filament towards the slits and the screen. Its quantum wave generally precedes it. The wave goes through both slits and produces an interference pattern after it has passed the slits. Each electron follows a well defined path, going through one slit or the other. It is now acted upon not only by the classical potential V but also the quantum potential Q . The quantum potential is derived from Schrödinger's equation and is expressed mathematically as

$$Q = -(\hbar^2 / 2m) [\text{second spatial derivative of } R] / R$$

2. Layers of brain organization relevant to consciousness
3. Protein conformational dynamics
4. Anesthesia and consciousness
5. Biological water
6. Quantum models
7. Conclusion

Quantum mechanisms in the brain?

1. Introduction

Are macroscopic quantum effects possible in the brain? As previous discussion has attempted to show, macroscopic quantum states in the brain would have great explanatory value toward understanding enigmatic features of consciousness. Macroscopic quantum states do exist in technological devices such as superconductors, and occur naturally as well. For example the core of a neutron star appears to be a gigantic Bose-Einstein condensate. Those who criticize the notion of relevant quantum states in the brain point out that quantum devices and quantum processes generally require temperatures near absolute zero to avoid thermal decoherence. While this is true for some quantum systems, this objection is not valid for all cases. For example superconductors operate at increasingly warmer temperatures as fabrication and efficiency improve (though technologically crafted devices are still nowhere near the precision, small scale and periodic perfection of some biological materials). Furthermore, quantum entanglement and EPR effects occur routinely at room temperature, in air or other media. And the core of a neutron star is exceedingly hot, not absolutely cold. Finally, biology is still mysterious. We don't really understand the 'living state'. We don't know what mechanisms for generating quantum states and avoiding thermal decoherence may have evolved over the billions of years of life on earth.

The issues of quantum isolation and avoidance of decoherence, evolution and relevance of proposed quantum mechanisms to cognitive processes will be discussed in later lectures. This week will cover neural and molecular correlates of consciousness, anesthesia, and how/where quantum mechanisms could beneficially operate in the brain.

Layers of brain organization relevant to consciousness

Attempts to approach consciousness often view the brain as a hierarchical system, comprised of layers of organization with bottom-up, as well as top-down feedback. In particular Alwyn Scott has elucidated such hierarchical organization, most notably in his book "Stairway to the Mind". In this view

consciousness emerges as a novel property at an upper level of the hierarchy from nonlinear interactions among layers.

The bottom level in Al Scott's scheme (and that of most others) is the membrane protein, implying ion channels and receptors as the fundamental units of information processing and representation. However this cutoff at the level of neurons or membrane proteins is arbitrary. Since consciousness is not understood we may need to go deeper. Furthermore the standard dogma in neuroscience---neurons or synapses as fundamental bits---is aimed at simplification to allow easier understanding, specifically to make the brain more like something with which we are familiar (e.g. a classical computer). These prevalent models are based on unfounded assumptions, as discussed below. Figure 1 illustrates a more complete hierarchical scheme.

Figure 1. Hierarchical levels of brain organization. Dotted line indicates cutoff for conventional approaches.

Since we can't measure or directly observe consciousness, how can we determine which levels/structures/activities may be related to consciousness?

We can ask in two ways: a) what are the levels/structures/activities seemingly most active in cognitive functions often associated with consciousness (attention, learning, perception, volition etc,---the 'neural correlate of consciousness'), and b) what structures/activities mediate the absence of consciousness (e.g. anesthesia).

At the systems level the current leading candidate for the "neural correlate" of consciousness involves neuronal circuits oscillating synchronously in thalamus and cerebral cortex. It has been known since the pioneering work of Herbert Jasper in the 1950's that sensory input passes through the thalamus where it is "broadcast" to cortex (e.g. the lateral geniculate nucleus in thalamus mediates visual information from the optic nerves; the information is then conveyed to visual cortex). Some thalamic-cortical projections carry specific sensory modalities whereas others are non-specific, but necessary for arousal and consciousness. The "reticular activating system" also passes through thalamus.

In recent decades numerous studies have also revealed extensive downward projections from cortex to thalamus. A consensus view has emerged in which reverberatory feedback ("recurrent loops") between thalamus and pyramidal cell neurons in cortex provides the "neural correlate of

consciousness" (e.g. Bernie Baars "global workspace"). Electrophysiological recordings further reveal coherent firing of thalamo-cortical loops with frequencies varying from slow EEG frequencies (2 - 12 Hz) to rapid gamma oscillations in the 40 Hz range and upward. Coherent gamma frequency thalamo-cortical oscillations (collectively known as "coherent 40 Hz") are suggested to mediate temporal binding of conscious experience (e.g. Singer et al 1990; Crick and Koch, 1990; Joliot et al, 1994; Gray, 1998). The proposals vary, for example as to whether coherence originates in thalamus or resonates in cortical networks, but reverberatory loops of "thalamo-cortical 40 Hz" activity stands as a prevalent view of the neural-level substrate for consciousness (e.g. Baars, 1988).

Figure 2. Left: Schematic of thalamocortical loops (from Newman, 1997). Right: Occurrence of high density of gap junctions in mammalian brain (Micevych and Abelson, 1991). Gap junctions are found throughout the brain, but in particularly high density in thalamus and cortex.

The prevalent view is that the neuronal network circuits involved in cognition and consciousness (e.g. thalamo-cortical loops) are selected and maintained through one particular type of neural-neural interaction: axonal-to-dendritic chemical synaptic transmission. The general idea is that the terminal axon of "Neuron A" releases neurotransmitter which then binds to a post-synaptic receptor on a dendritic spine on "Neuron B". This changes Neuron B's local dendritic membrane potential which interacts with those on neighboring spines and dendrites on Neuron B and can reach threshold to trigger axonal depolarizations ("firings", or "spikes") down Neuron B's axon. These travel along the axon and result in release of neurotransmitter vesicles from Neuron B's pre-synaptic axon terminal into another synapse. The neurotransmitter then binds to post-synaptic dendritic spine receptors on Neuron C, and so on. The basic idea is that in each neuron multiple inputs and analog processing in dendrites reaches an "all or none" threshold for axonal firings—a single, digital output seen as the fundamental 'bits', or units of information in the brain.

Figure 3. "Pyramidal cell" cortical neurons are considered to be essential neurons for higher processes in primates, and likely sites for conscious processes

Figure 4. Pyramidal cells form axonal-dendritic synapses. Dendritic-dendritic connections are seen between cells at bottom.

Particular landscapes of chemical synaptic network patterns, presumed to correlate with a particular mental state, arise and change dynamically in this view by the relative strengths of the chemical synapses. The synaptic strengths shape the network dynamics and are responsible for cognitive effects such as learning by selecting particular neural network patterns (e.g. Hebb, 1949).

However there are other possibilities for information processing among neurons. Both the late Sir John Eccles and Karl Pribram have advocated dendritic-dendritic processing, including processing among dendrites on the same neuron. Eccles in particular pointed at dendritic arborizations of pyramidal cells in cerebral cortex as the locus of conscious processes. Francis Crick suggested that mechanical dynamics of the dendritic spines were essential to higher processes including consciousness.

Another "fly in the ointment" of the conventional approach is the prevalence of gap junction connections in the brain. In addition to chemical synaptic connections, neurons (and glia) are widely interconnected by electrotonic gap junctions. These are window-like portholes between adjacent neural processes (axon-dendrite, dendrite-dendrite, dendrite-glial cell). Cytoplasm flows through the gap, which is only 4 nanometers between the two processes, and the cells are synchronously coupled electrically via their common gap junction. Eric Kandel remarks that neurons connected by gap junctions behave like "one giant neuron". The role of gap junctions in thalamo-cortical 40 Hz is unclear, but currently under investigation.

Figure 5. A gap junction is a window between adjacent cells through which ions, current and cytoplasm can flow.

Gap junctions are generally considered to be more primitive connections than chemical synapses, essential for embryological development but fading into the background in mature brains. However brain gap junctions remain active throughout adult life, and are being appreciated as more and more prevalent (though still sparser than chemical synapses - a rough estimate is that gap junctions comprise 15% of all inter-neuron brain connections). Gap junctions may be important for macroscopic spread of quantum states among neurons (and glia). Biological electron tunneling can occur up to a distance of 5 nanometers, so the 4 nanometer separation afforded by gap junctions may enable tunneling through the gaps, spreading the quantum state. In the past few years specific intracellular organelles have been discovered in dendrites, immediately adjacent to gap junctions. These are layers of membrane covering (at least in some cases) a mitochondrion, and are called "dendritic lamellar bodies - DLBs" (de Zeeuw, 1995). The DLBs are tethered to small cytoskeletal proteins anchored to microtubules. Perhaps mitochondria provide free electrons for tunneling, and DLBs form a sort of Josephson junction between cells, permitting spread of cytoplasmic quantum states throughout widespread brain regions?

Figure 6. Dendritic lamellar bodies (DLBs) are situated on opposite sides of gap junctions in brain neuronal dendrites. Are they quantum tunneling devices spreading a unitary quantum state among neurons?

In the standard model axonal firings resulting in neurotransmitter release are the basic currency of information, however only a fraction of them (~15%) result in neurotransmitter release. The probability of release seems random, although local factors (?cytoskeleton) could be modulating the release. Beck and Eccles (1992) suggest that quantum effects are influencing the process. Much of what we know about neuronal behavior within the brain has been derived from electrophysiological recordings. Typically such recording

includes a great deal of background drift, or "noise", which electrophysiologists commonly eliminate by taking the average of many recordings. But how do we know that the background drift in individual neurons or local regions is indeed noise? Israeli neurophysiologist Aviram Grinvald asked that question and recorded "background noise" (he terms it "ongoing activity") simultaneously in various areas of mammalian brain. Grinvald found that the ongoing activity correlates across brain regions! Could it represent some heretofore unrecognized signaling or information? If so, how? Perhaps electrophysiological averaging is "throwing away the baby with the bathwater"? What about activity within individual neurons? Could Grinvald's "ongoing activity" stem from some internal drive? Is the neuron truly the fundamental unit of information, or are processes relevant to consciousness going on within cytoplasm, and/or at the molecular, and quantum levels?

Figure 7. Connections between membrane/synaptic events and cytoskeleton.

Figure 8. Dendritic spine glutamate receptors connect to cytoskeletal components including microtubules. From Von Bissen et al Trends in Neuroscience July, 1999

Synapses and membrane proteins are considered in conventional approaches as the bottom rung in the hierarchical arrangement leading to cognition and consciousness. But membrane proteins are short-lived, and turn over on a scale of hours to days. They are maintained by a system of axoplasmic transport, in which membrane proteins and other materials necessary for synaptic functions are synthesized in the cell body and supplied to the synapse by microtubule-dependent transport (aided by feedback signals from synapse to cell body). There is some debate as to whether the transport utilizes microtubules in purely a passive manner (the 'rails' on which vesicles and other materials move) or actively direct the materials. In any case the microtubules and other cytoskeletal structures also establish new synapses, and up-regulate and down-regulate existing synaptic efficacy. So

the sensitivity of synapses---the cornerstone of learning and cognitive function---depends on cytoskeletal microtubules. Microtubules within neurons may also communicate among dendrites, and propagate "error" signals (as in neural network "back-propagation") to facilitate learning.

Figure 9

Materials for synapses are synthesized in cell body, and transported along microtubules.

Do neurons utilize intracellular information processing? Assuredly they do, however conventional approaches assume this occurs by diffusion of soluble biomolecules (second messengers, ionic flux etc). Perhaps, but the cytoskeleton provides a solid-state system well suited for rapid information processing (and proven to convey information, such as in Ingber's experiments) whereas diffusion of biomolecules (especially across great distances through neural processes) seems more like a minestrone soup. Single cell organisms such as paramecium swim, avoid obstacles, find food, mates and have sex. As single cells they have no synapses---they utilize membrane and especially cytoskeletal microtubules as their nervous system. If a single cell paramecium can swim around and find food and hot dates, isn't a single cell neuron better than a simple on-off switch? Processing within the neuron, for example in the cytoskeleton might well be relevant to consciousness.

Figure 10. Paramecium "orgy". Mating behavior of unicellular paramecia involve pairs (or more) of the creatures fusing cytoplasm, becoming "one" organism. Paramecia are absolutely still during mating. This behavior is accomplished without synapses, as paramecia are single cell organisms.

3. Protein conformational dynamics

In terms of real-time dynamical regulation of cellular activity the conformational states of proteins are the most important factor. This might include an ion channel opening or closing, a receptor changing shape upon binding of neurotransmitter, an enzyme catalyzing a reaction, for a "second

messenger" or cytoskeletal protein changing shape to signal or facilitate movement or transport. How are protein conformational states regulated? Protein function depends on shape, or conformation. Individual proteins are synthesized as linear chains of amino acids which "fold" into 3 dimensional conformation. The precise folding depends on attractive and repellent forces among various amino acid side groups, and a current view is that many possible intermediate conformations precede the final one (Baldwin, 1994). Predicting final 3-dimensional folded shape using computer simulation has proven difficult if not impossible. This conundrum is known as the "protein folding problem" and so far appears to be "NP complete": the answer can be calculated in theory, but the space and time required of any classical computer is prohibitive. Perhaps protein folding is a quantum computation? (Crowell, 1996).

The main driving force in protein folding occurs as uncharged non-polar amino acid groups join together, repelled by solvent water. These "hydrophobic" groups attract each other by dipole couplings known as van der Waals forces and bury themselves within the protein interior. Intra-protein "hydrophobic pockets" result, composed of side groups of non-polar (but polarizable) amino acids such as leucine, isoleucine, phenylalanine, tryptophan, tyrosine and valine. Volumes of the hydrophobic pockets (~400 cubic angstroms, or 0.4 cubic nanometers) are roughly 1/30 to 1/250 the total volume of a single protein, and their physical solvent characteristics most closely resemble olive oil (e.g. Franks and Lieb, 1985). Van der Waals forces in hydrophobic pockets establish protein shape during folding, and also regulate dynamic conformational changes.

Proteins in a living state are dynamical, with transitions occurring at many scales, however conformational transitions in which proteins move globally and upon which protein function generally depends occur in the nanosecond (10^{-9} sec) to 10 picosecond (10^{-11} sec) time scale (Karplus and McCammon, 1983). Proteins are also only marginally stable. A protein of 100 amino acids is stable against denaturation by only ~40 kilojoules per mole (kJ mol^{-1}) whereas thousands of kJ mol^{-1} are available in a protein from side group interactions including van der Waals forces. Consequently protein conformation is a "delicate balance among powerful countervailing forces" (Voet and Voet, 1995).

The types of forces operating among amino acid side groups within a protein include charged interactions such as ionic forces and hydrogen bonds, as well as interactions between dipoles---separated charges in electrically

neutral groups. Dipole-dipole interactions are known as van der Waals forces and include three types:

- 1) Permanent dipole - permanent dipole
- 2) Permanent dipole - induced dipole
- 3) Induced dipole - induced dipole

Figure 11. London forces

Type 3 induced dipole - induced dipole interactions are the weakest but most purely non-polar. They are known as London dispersion forces, and although quite delicate (40 times weaker than hydrogen bonds) are numerous and influential. The London force attraction between any two atoms is usually less than a few kilojoules, however thousands occur in each protein. As other forces cancel out, London forces in hydrophobic pockets can govern protein conformational states.

Figure 12. Schematized protein capable of switching between two conformational states governed by van der Waals interactions in a hydrophobic pocket. Proteins may actually have several smaller collectively governing hydrophobic pockets. Top: Protein switching between 2 conformational states coupled to localization of paired electrons (London force) within a hydrophobic pocket. Bottom: quantum superposition (simultaneous existence in two distinct states) of the electron pair and protein conformation.

London forces ensue from the fact that atoms and molecules which are electrically neutral and spherically symmetrical nevertheless have instantaneous electric dipoles due to asymmetry in their electron distribution. The electric field from each fluctuating dipole couples to others in electron clouds of adjacent non-polar amino acid side groups. Due to inherent uncertainty in electron localization, London forces are quantum effects which may couple to "zero point fluctuations" of the quantum vacuum (London, 1937; Milloni, 1994).

Quantum dipole oscillations within hydrophobic pockets were first proposed by Frohlich (1968) to regulate protein conformation and engage in macroscopic coherence, and Conrad (1994) suggested quantum superposition of various possible protein conformations occur before one is selected. Roitberg et al (1995) showed functional protein vibrations which depend on quantum effects centered in two hydrophobic phenylalanine residues, and Tejada et al (1996) have evidence to suggest quantum coherent states exist in the protein ferritin.

Evidence for a pivotal role for quantum effects in protein conformational regulation (and consciousness) comes from studying the opposite of consciousness - anesthesia.

Anesthesia and consciousness

Another clue to the microsite of consciousness is the molecular mechanism of anesthesia. At just the right anesthetic dose, consciousness is erased while other brain activities continue. Understanding the precise mechanism of anesthesia may illuminate consciousness, but first let's consider the interesting history of anesthesia (the following is excerpted from an article I wrote in a forthcoming book entitled "Greatest Inventions of the Past 2000 Years" edited by Sara Lippincott and John Brockman, Simon and Schuster). Anesthesia grew from humble beginnings. Inca shamans performing trephinations (drilling holes in patients' skulls to let out evil humors) chewed coca leaves and spat into the wound, effecting local anesthesia. The systemic effects of cocaine were studied by Sigmund Freud, but cocaine's use as a local anesthetic in surgery is credited to Austrian ophthalmologist Karl Koller who in 1884 used liquid cocaine to temporarily numb the eye. Since then dozens of local anesthetic compounds have been developed and utilized in liquid solution to temporarily block nerve conduction from peripheral nerves and/or spinal cord. The local anesthetic molecules bind specifically on sodium channel proteins in axonal membranes of neurons near the injection site, with essentially no systemic effects on the brain. On the other hand general anesthetic molecules are gases which do act on the brain in a remarkable fashion - the phenomenon of consciousness is erased completely while other brain activities continue.

General anesthesia by inhalation developed in the 1840's, involving two gases used previously as intoxicants. Soporific effects of diethyl ether ("sweet vitriol") had been known since the 14th century, and nitrous oxide ("laughing gas") was synthesized by Joseph Priestley in 1772. In 1842 Crawford Long, a Georgia physician with apparent personal knowledge of "ether frolics" successfully administered diethyl ether to James W. Venable

for removal of a neck tumor. However Long's success was not widely recognized, and it fell to dentist Horace Wells to publicly demonstrate the use of inhaled nitrous oxide for tooth extraction at the Massachusetts General Hospital in 1844. Although Wells had apparently used the technique previously with complete success, during the public demonstration the gas-containing bag was removed too soon and the patient cried out in pain. Wells was denounced as a fake, however two years later in 1846 another dentist William T.G. Morton returned to the "Mass General" and successfully used diethyl ether on patient William Abbott. Morton used the term "letheon" for his then-secret gas, but was persuaded by Boston physician/anatomist Oliver Wendell Holmes (father of the Supreme Court Justice) to use the term anesthesia.

Figure 13 William T.G. Morton administering anesthesia to William Abbott at Massachusetts General Hospital in 1846.

Modern anesthesia is safe and effective, however one major mystery remains. Exactly how do anesthetic gases work? The answer may well illuminate the grand mystery of consciousness. Inhaled anesthetic gas molecules travel through the lungs and blood to the brain. Barely soluble in water/blood, anesthetics are highly soluble in a particular lipid-like environment akin to olive oil.

In 1897 Meyer in Germany and Overton in England independently came to the same conclusion: over many orders of magnitude potency of the various anesthetic gases was directly proportional to the gas solubility in a particular lipid-like environment, most closely identified with olive oil.

Figure 14. The Meyer-Overton correlation. Anesthetic potency plotted against solubility in olive oil for 13 gas anesthetics yields a straight line over 70,000 fold increase. MAC (inverse of potency) is the mean alveolar concentration (in atmospheres) of anesthetic gas concentration measured in lung alveoli in equilibrium with brain concentration, for which 50% of subjects will respond to painful stimuli.

It turns out the brain is loaded with such oily stuff, both in lipid membranes and tiny water-free ("hydrophobic") lipid-like pockets within certain brain proteins. For many years it was assumed that anesthetics acted in lipid regions of membranes, however Nicholas Franks and William Lieb at Imperial College in London showed in a series of experiments during the 1980's that anesthetics inhibit individual protein function even when there is no membrane in the picture. They concluded that anesthetics exert their effect by entering lipid-like ("hydrophobic") pockets within proteins and forming very weak van der Waals bonds. The hydrophobic pockets are typically less than 1/30 of the total protein volume, but seem to somehow act as the "brain" of the protein. Somehow weak (quantum) interactions in these tiny regions have profound effects on the function of particular proteins, and consciousness.

Figure 15. Computer simulation of the anesthetic-sensitive enzyme papain with the anesthetic gas molecule halothane (hatched) "docked" by energy minimization into its major hydrophobic pocket (black). Scale bar: 1 nanometer. (From Louria and Hameroff, 1996 with permission).

Which brain proteins mediate anesthetic effect? Sites for immobility during anesthesia are in spinal cord and those for amnesia are largely in hippocampus (Eger et al, 1996), however sites essential for anesthetic erasure of consciousness appear to be diffusely located, particularly in thalamo-cortical projections.

Franks and Lieb (1998) conclude that the particular proteins most sensitive to inhalation anesthetics are post-synaptic receptors for GABAA, glycine, serotonin 5HT3, and nicotinic acetylcholine as well as some potassium channels. Franks and Lieb add that other proteins (e.g. voltage-sensitive channels, cytoskeletal actin and microtubules etc.) which are less anesthetic-sensitive but more abundant and/or directly involved in activities relevant to consciousness may also mediate anesthetic effects. Recent evidence implicates G-proteins and gap junction proteins in anesthetic mechanism. Normal function of glycine receptors (Delon and Legendre, 1995) and of GABAA receptors (Whatley et al, 1994) depend on the integrity of cytoskeletal microtubules, so it seems likely that a variety of receptors, channels, second messenger and cytoskeletal proteins engage in collective dynamics necessary for consciousness (and inhibition by anesthetics). This

can explain why disruption of either excitatory (e.g. acetylcholine) or inhibitory (e.g. GABAA) receptor function contribute to anesthesia. The essential feature common to molecular sites of anesthetic action is the hydrophobic pocket.

Franks and Lieb (1994) had concluded that anesthetics act simply by following the Meyer-Overton correlation: their mere presence in hydrophobic pockets prevents conformational switching. However a variety of molecules which follow the Meyer-Overton correlation and occupy the same hydrophobic pockets are nonanesthetic, or even convulsant (Fang et al, 1996). The mere presence of molecules in hydrophobic pockets is insufficient to explain anesthesia.

A logical conclusion is that anesthetics somehow disrupt van der Waals London force interactions normally occurring in the critical hydrophobic pockets. Quantum superposition (theoretically implicated in consciousness) requires electron mobility---electron pairs must be relatively free to roam among allowed orbitals. Evidence shows that anesthetics retard electron mobility---the movement of free electrons in a corona discharge is inhibited by anesthetics (Hameroff and Watt, 1983). By forming their own London force attractions in hydrophobic pockets, anesthetics may retard electron mobility required for protein dynamics, quantum superposition and consciousness. Nonanesthetics may be understood as occupying hydrophobic pockets without altering electron mobility, and convulsants as forming cooperative van der Waals interactions which promote excessive electron mobility and protein dynamics in certain brain proteins.

Figure 16. A. Schematic of anesthetic (A) in hydrophobic pocket retarding electron mobility thus preventing protein conformational change and quantum coherent superposition. B. Convulsant molecule (C) in hydrophobic pocket promotes electron mobility and protein dynamical switching.

Can we learn about consciousness from drugs which have opposite effects from anesthetics? In the 1970's pharmacologists Jack Green, Solomon Snyder and others studied a series of drugs with psychedelic, or hallucinogenic properties. Much like Meyer and Overton had ranked anesthetic potency and found a physical correlation with solubility in olive oil, Green, Snyder and colleagues looked for physical correlates of the drugs'

hallucinogenic potency. They found that potency correlated with the degree of electron resonance energy donation from drug molecule to receptor. The more electron energy able to be transferred, the greater the hallucinogenic potency. Such electron energy donation/mobility would seem to enhance likelihood of quantum superposition states in receptor proteins, the exact opposite of anesthetics. Altered ("psychedelic") states may involve enhanced quantum superpositions, equivalent to expanded access to normally non-conscious quantum information.

Figure 17 A. Anesthetic gas molecule (A) in a hydrophobic pocket of critical brain protein (receptors, channels, tubulin etc.) prevents normally occurring London forces, protein conformational dynamics and superposition necessary for consciousness. B. A psychedelic hallucinogen (P) acts in hydrophobic pocket in critical brain protein to promote and sustain superposition, 'expanding' consciousness.

See also Anesthesia, consciousness and hydrophobic pockets-a unitary quantum hypothesis of anesthetic action. Toxicology Letters 100/101:31-39 (1998). ~<http://www.u.arizona.edu/~hameroff/calgary.html>
<<http://www.u.arizona.edu/hameroff/calgary.html>>

5. Biological water

An older theory of anesthesia (e.g. put forward by Linus Pauling) involved the protein-water interface, and the formation of what Pauling called "clathrates", an abnormal form of water. We now know that water molecules form all kinds of transient forms, and also that protein conformation mediated by hydrophobic pockets influences binding at protein surfaces (Wulf and Featherstone, 1967). So the state and ordering of water in biological systems depends on the states of proteins and other surfaces.

Water adjacent to surfaces of cytoskeleton and membranes may be ordered by those structures, leading to layers of ordered water within cells. As cytoskeletal structures may be quite dense, particularly in actin gelation, its possible that nearly all water within a cell or region of cell may be ordered. There is a long historical story proposing ordered water leading to biological quantum states stemming from Stuart, Ricciardi, Umezawa, Takahashi, del Giudice, Vitiello and more recently Jibu, Yasue and Scott Hagan. We'll hear about this in some detail next week from Scott Hagan. In one aspect of this approach, Jibu, Yasue, Hagan and others have proposed that ordered water

in the hollow microtubule core is ordered by microtubule dynamics leading to spontaneous symmetry breaking and generation of evanescent photons, "super-radiance" and "self-induced transparency" of photons in MT cores. This is illustrated in Figure 18.

Figure 18 Jibu, Yasue and Hagan Super-radiance in Microtubules (from Jibu et al, 1994). A schematic representation of the process of super-radiance in a microtubule. Each oval without an arrow stands for water molecule in the lowest rotational energy state. Each oval with an arrow stands for a water molecule in the first excited rotational energy state. The process is cyclic (a b c d a b . . .), and so on. (a) Initial state of the system of water molecules in a microtubule. Energy gain due to the thermal fluctuation of tubulins increases the number of water molecules in the first excited rotational energy state. (b) A collective mode of the system of water molecules in rotationally excited states. A long-range coherence is achieved inside a microtubule by means of spontaneous symmetry breaking. (c) A collective mode of the system of water molecules in rotationally excited states loses its energy collectively, and creates coherent photons in the quantized electromagnetic field inside a microtubule. (d) Water molecules, having lost their first excited rotational energies by super-radiance, start again to gain energy from the thermal fluctuations of tubulins, and the system of water molecules recover the initial state (a). spot, and initiated movement. Albrecht-Buehler further showed that the light-sensitive organelle perceiving direction is the microtubule-based centriole (a pair of microtubule mega-cylinders) within the cell near the nucleus. Jibu and Yasue [2] have argued that the electromagnetic signaling from the distant light spot reaching the centriole consists of "evanescent" photons tunneling through dynamically ordered regions of cell water.

Also, charged surfaces (cytoskeleton, membranes) attract soluble ions of the opposite charge, leading to plasma-like layers (Debye layer) which can have quantum properties, and/or quantum isolation properties. Such plasma-like regions have been proposed adjacent to membranes (Triffet and Green), and surrounding microtubules. Dan Sackett at NIH has shown that at physiological pH there exists a plasma sleeve around microtubules which could serve to isolate MT quantum processes from thermal decoherence.

Figure 19 Dan Sackett at NIH has shown that at proper physiological pH microtubules are surrounded by a plasma phase of ions and counter ions. The terminal chain of amino acids of each tubulin sticks out like a hair with charges which attract oppositely charged ions, resulting in a plasma layer which could house quantum effects, and/or shield the MT from environmental decoherence.

6. Quantum models

Below is a partial list of quantum models with specific proposed biological structures and quantum mechanisms pertaining to consciousness. In Week 7 we will dissect these approaches and consider issues pertaining to thermal decoherence and isolation.

Neuronal synapses - electron tunneling - Walker, 1970

Neuronal cell water - quantum field ordering - Stuart, Takahashi, Umezawa, 1978

Neuronal firings - superposition/objective reduction - Penrose, 1989

Neural membrane/synaptic proteins - Bose-Einstein condensate - Marshall, 1989

Pre-synaptic vesicle release - quantum indeterminacy - Beck/Eccles, 1992

Calcium ions causing vesicle release - wave function collapse - Stapp, 1994

Neural proteins - quantum superposition/interference - Conrad, 1992

Microtubules - quantum coherence - Hameroff, 1994

Ordered water at protein surfaces - evanescent photons - Jibu/Yasue, 1994

Ordered water within microtubule cores - super-radiance/self-induced transparency - Jibu/Yasue, 1994

Microtubules - quantum computation/objective reduction 'Orch OR' - Penrose/Hameroff, 1995, 1996

Lipids in neural membrane e.g. - quantum interference - Wallace, 1996

Gap junctions - electron tunneling - Hameroff, 1998

7. Conclusion

Conventional approaches to consciousness stop at the level of individual cells, or membrane proteins/synapses. These approaches fail to explain enigmatic features related to consciousness. Consequently it is prudent to look deeper, into the neuron and down to the quantum level. Quantum mechanisms have explanatory power, however there are seemingly difficult barriers in terms of understanding how quantum mechanisms could survive

in the warm, apparently noisy environment of the brain. These issues will be discussed next week.

Protein tactilization/percolation, electron superposition/conformation and the Born-Oppenheimer approximation

Michael Conrad

Quantum processes could contribute to biological information processing without the requirement for coherent states. The idea is that conformational-electronic interactions in proteins and other biological macromolecules allow for a close classical-nonclassical interface. Superpositions of electronic states then speed up molecular recognition processes. This in turn enhances cellular (including neural) recognition-action capabilities, and consequently the perception action capabilities of the organism. The quantum superpositions in effect percolate up to the macro level of organization. We have a vertical (cross-scale) model of information processing. Information is processed at each level: the intercellular, the intracellular, and the molecular (including electronic-conformational interactions). But the molecular level is intrinsically the most powerful, due to the powerful ability of proteins (in particular) to discriminate particular object molecules in a complex environment.

I refer to the above as the quantum speedup principle. The superpositional parallelism of the electronic wave function is in effect converted to speed of enzyme recognition (or to enhanced recognition capability). The quantum speedup principle is not alien to a wet, warm brain. It utilizes this feature, since it essentially funnels thermal energy through electronic superpositions to selected degrees of freedom of the nuclear coordinates.

Let us back up a bit and explain this process more clearly, though still informally (for technical details see references below). In ordinary organic molecules the electrons, being low mass particles, follow the nuclear motion very closely. For the purposes of analyzing the infrared spectrum the electrons are essentially wrapped together with the atomic nuclei. This is the Born-Oppenheimer approximation.

But in proteins and other macromolecules it is now known that there are loosely bound electrons, for example electrons that tunnel significant distances through hydrogen bond pathways and also various surface electrons. (A typical protein contains about 18,000 electrons. To this must be added the various electrons in the surround, for example in the sphere of hydration that envelops the protein or in the polar heads of the lipid

membrane components, if we happen to be dealing with a protein that swims in the membrane).

These loosely bound electrons will undergo acceleration relative to the nuclei. Thus they must emit photons (in the infrared) that cannot be attributed to the motions of the nuclear charges. It means that the electrons will be in closely spaced energy states, due to interaction with the ubiquitous infrared radiation field. They will continually be falling to lower states and continually absorbing photons and raised to higher states. The more closely spaced the energy states the longer lived the superposition (in accordance with the time-energy uncertainty principle). The important point is that we expect superpositions, and therefore we expect time dependent interference effects. Each eigenfunction contributing to the superposition vibrates with a frequency that depends on the energy, in accordance with $E = h\nu$. Electrons in a superposition of two or more energy states will exhibit a superposition of vibration frequencies, hence the probability that they will be found in any particular place in space will change with time (due to the changing phase relations among the different contributing eigenfunctions). The partial charge will thus be time dependent (this is the average charge in any particular region). The variation in the partial charge can agitate hydrogen bonds (protons), in particular mobile protons that run in chains through proteins and nucleic acids. They can even affect the motions of heavier nuclei, due to the fact that it is the effective mass (not the actual mass) that is the pertinent factor. The nuclei are bound in a structure. It may be very hard to push them in some particular direction, because of the forces acting on them. The effective mass is higher than the actual mass in these directions. But on the other side there are other directions in which a small push can put the nuclei under the influences of forces that move it further in that direction. The effective mass will be low in those directions. Any agitation resulting from funneling thermal energy (energy associated with the infrared radiation field) through superpositions to motions of nuclei increases the dynamics of the protein, the chance of making conformational transitions, and does so in a selective way.

Let us step back even a bit further and look at how one can think of protein recognition. The lock-key metaphor is often used. The protein is like a key that recognizes the substrate in lock-key fashion. (Actually this metaphor was developed by Emil Fisher, who thought of the enzyme as the lock, since enzymes are usually bigger than the metabolites they act on. But for our purposes here I think it is more graphic to think of the enzyme as the key, since it is the more active entity.)

The lock-key metaphor (or a jigsaw puzzle metaphor if we are thinking of self-assembly of macromolecules to form higher level structures) is fine so far as understanding why two molecules fit together. Many weak interactions contribute. Perhaps the most pertinent are the London dispersion forces (transient dipole-induced dipole) since these fall off as the 7th power of distance and are universal. So if molecules are big enough to have surface properties they can recognize each other, provided that at the same time they are small enough to explore each other's shapes through Brownian search. The conformational-electronic interaction described above makes this Brownian search more effective.

There is a flaw in the above metaphor, connected with the idea of search. What is the chance that a key will randomly find a keyhole, especially if the fit is very specific? This is the docking problem. Docking requires dynamics. Sometimes the lock-key analogy is replaced by a more dynamic hand-glove analogy (proposed by Koshland). The important point is that enzymes appear to draw in the substrate. One would think that this would make it more difficult for them to eject the substrate and go on to another reaction. But this effect is very minor, at most. The conformational-electronic picture is pertinent to docking. It suggests that biological macromolecules are in perennial ordered motions (corresponding to the picture from nuclear magnetic resonance, though of course this does not tell whether the motions are in any way ordered). The enzyme can actively draw in and eject the substrate, since the electronic superposition essentially serves to create instability. What we want is really instability that will facilitate folding, self-assembly, or complex decomposition (in the case of enzyme action).

Let us step back one step further and look at two views as to how biomolecular systems can be treated. One is the molecular dynamics view. The idea is that quantum mechanics is to be used to ascertain bond strengths between nuclei, but that once this is done one can proceed classically, essentially by solving Newton's equation for the time development. Quantum features are buried at such a level that they are irrelevant, just like the forces within the atomic nuclei are presumably irrelevant (since the nuclei are dormant at biological temperatures). The view that emerges from analysis of the conformational-electronic interaction is that quantum features are pertinent to time development, at the level of the macromolecular interaction and consequently at the level of cells and organisms. The quantum features percolate up into the time development of the system.

What does this mean for brain capabilities? Take the binding problem. Cells must respond appropriately to electrochemical input sequences from other cells, or to modulating influences of various sorts. What puts a time order on this? Certainly not the length of the connections in the brain. There are too many. As far as we know there is no fast clock in the brain, though there is some speculation about particular frequencies being important. The model presented here opens up another possibility. We think of cells (including neurons) as powerful pattern processors, able to disambiguate (or appropriately group) spatially ambiguous input patterns. The cell uses its pattern processing, which is largely derived from the quantum speedup principle, to chunk the inputs continually coming to it in a meaningful way. In order to do this one needs a fast signal processing system in the cell. The cytoskeleton is likely to serve this function.

Let us step back one more step and look at the pattern recognition problem. It is combinatorially explosive. If you have an n bit pattern and want to classify it in two groups we have (2^n) groupings. No machine can make much of an impact on this; nor is it reasonable to suppose that organisms can solve the problem. But organisms can use the multiple interactions available to them to do a lot better than programmable machines. After all, with n particles we have n^2 potential interactions. When the quantum effects are taken into account we have interference among possible locations of the electrons, which probably grows exponentially (according to Feynman).

Let us step forward now, and look at the relevance to consciousness. Is there any? As indicated in my paper, Principle of Philosophical Relativity, I am not inclined to believe that interpretations of consciousness that will meet with the kind of universal acceptance that empirical or mathematical scientific discoveries meet with is likely. But it is important that the theories of science be eligible for interpretations in terms of consciousness. Here we have quantum superpositions (qualities), percolation of superpositions (presumably with some sort of collapse process) and therefore something akin to decision making, and we have probabilistic features (so at least we do not exclude free will). And all this travels automatically with increased perception-action capabilities.

Of course to analyze this we have to move on to brain architectures that allow percolation of micro influences to the macro level, and also the influence of macro level inputs in meso and micro scale events.

There are other issues that come up in connection with the model described above, which are more fundamental. The electronic transitions are accelerations. According to general relativity acceleration is identical to

gravity, hence a nonlinear phenomenon. The linear superposition principle cannot hold. So there should as a whole be a connection to wave function collapse. This point is addressed in the fluctuon model (briefly described in the Principle of Philosophical Relativity Paper) and is the broader background of the model presented here.

One further point. What is the connection to coherence? We don't need coherence of any kind for the quantum speedup principle. Coherence can refer to many things. The quantum speedup principle does not exclude the possible importance of coherence for the organism. And it may be possible for coherent effects to occur at body temperatures. I can perhaps here draw attention to a model I developed some years ago, which might be called the proton supermobility model. In an ordinary superconductor electrons polarize nuclei, which attract other electrons, leading to an attractive interaction between electrons (described as a phonon interaction in field theoretic terms). We reverse the picture at the water-membrane interface. Mobile protons migrating in the layer of bound water adjacent to the membrane and enveloping the macromolecular components of the cell polarize electrons in the polar head groups of membrane lipids, thereby leading different proteins to be attracted to the same charge distortion and therefore to each other. The problem with the model is this. To have coherence the de Broglie wave lengths should overlap. It takes a reduction of from 4 to 7 times in effective mass of the protons in water for this to happen. But it is possible that the effective mass of the protons is altered by macromolecular interactions and therefore that superfluid threads of protons could develop, allowing for a dynamic order in the cell (a thought that was attractive to London, Schrodinger, and other early thinkers in the field of low temperature physics).

Such a rigid but dynamic thread-like substratum would provide a kind of hidden skeleton for the cell and would allow the cell to react in a specific way to influences impinging on different regions. The threads could run from cell to cell. But the model, and any model of this type, should be taken with a grain of salt at this stage. Experimentation in this area is very difficult, subject to artifacts. Water, the bath of life, is terribly difficult to study. The effects may occur, but they have to be strong enough to actually play a significant role. Nevertheless I wanted to bring up this point, since it helps to delineate the quantum speedup principle (which on fundamental grounds seems sound) from models involving coherence. They are different ways in which quantum mechanics can play a role in biological systems, but

whether they do so and the extent of the contribution are important open questions.

Feasibility of macroscopic quantum mechanisms in the brain

Stuart Hameroff

Lecture 1

1. Introduction

Quantum claims: Levels of audacity

Assumptions: Is the brain "warm, wet and noisy?"

Summary of the Orch OR model

Feasibility of quantum states in the brain: reply to quantum critics Stegmark,

Tuszynski and Scott

Conclusion

1. Introduction

The conventional wisdom in academic science holds decidedly against the relevance of quantum mechanisms in consciousness (or living processes in general) above the level of inter-atomic, intra-molecular and inter-molecular forces which mediate chemical interactions. Above this level, the wisdom holds, quantum effects are washed out by thermal decoherence, or environmental interaction. Nonetheless, quantum effects above this level - macroscopic quantum states, or coherent states - have been suggested as solutions to problems in consciousness and living systems. These suggestions have engendered a rash of criticism. This week's topic is devoted to these issues - are macroscopic quantum states even feasible in biological systems such as the brain? How could environmental decoherence be avoided in a warm, wet, seemingly noisy environment? How could relevant quantum states be detected or verified?

2. Quantum claims: Levels of audacity

To begin, let's look at the types of claims or suggestions for quantum states in biological systems relevant to consciousness. They are presented here roughly in order of increasingly "strong" claims (or audacity).

Quantum effects mediate bonds, inter-atomic forces in biomolecules. This is generally agreed upon; there are no implications for macroscopic effects or coherence, nor for consciousness or the living state.

Quantum superposition originating in electrons and/or protons mediate functional activities in biomolecules, e.g. protein conformation (e.g. Michael Conrad, evidence from protein chemistry, protein folding problem) This does not necessarily imply cell-wide, or tissue-wide macroscopic effects or coherence but raises the scale of the quantum state to that capable of regulating a single protein's conformation.

Quantum influence on protein assemblies (e.g. pre-synaptic vesicular grid) releasing neurotransmitters into synaptic cleft (Beck and Eccles, 1992). This does not necessarily imply macroscopic or coherent effects, in fact Beck and Eccles saw the quantum influence as a purely random, probabilistic effect. But it suggests quantum states/influence on the scale of an organelle, or assembly of proteins.

Sites where anesthetics bind indicate the site, or at least partial site for consciousness.

It is a distributed set of micro-sites of hydrophobic pockets in various brain proteins which have in common a particular solubility parameter value (i.e. olive oil). Each tiny, oily pocket in the array is characterized by being conducive to

electron delocalization within the confines of the pocket. Electron dipole couplings in the tiny pockets control each protein's conformation and are quantum mechanical van der Waals London forces.

The anesthesia story thus indicates that protein conformational states, and consciousness, depend on quantum forces in hydrophobic pockets of certain brain proteins. This does not necessarily imply macroscopic or coherent effects, however such effects could be supported by the London forces (e.g. as a Bose-Einstein condensate).

Quantum influence on locations of ions, e.g. calcium ions, which influence neural function (e.g. Stapp). Macroscopic coherence implied through unity of quantum wave function

Bose-Einstein condensate among distributed neural proteins, e.g. using Frohlich mechanism to drive coherent excitations/phonons (Marshall, 1989). Macroscopic condensate manifests unitary consciousness. Marshall's distributed neural proteins may be the array of hydrophobic pockets which mediate anesthesia.

Macroscopic quantum state arise from ordered water effects, spontaneous symmetry breaking, super-radiance, generation of evanescent photons with 50 micron coherence length (Jibu, Yasue, Hagan)

Macroscopic quantum states in protein assemblies, in particular microtubules. The crystal-like lattice periodicity in microtubule structure facilitates tunneling/entanglement/condensation among quantum states in the arrayed subunit proteins (tubulin), with coherent pumping due to a Frohlich type mechanism, stochastic quantum resonance, or what Conrad describes as funneling.

The microtubule based quantum state spreads among different neuronal dendrites via tunneling through gap junctions, enabling brain-wide quantum states.

Quantum states of individual tubulins in microtubules function as 'qubits', and microtubules perform quantum computation. The quantum superposition/quantum computation phase is sustained long enough (10 - 500 msec) to reach threshold for self-collapse by Penrose's quantum gravity 'objective reduction'. Such objective reductions constitute conscious moments.

Critics/skeptics of the quantum approach will take issue with items 2 - 10. To anticipate their objections, in a following section I'll refer to 3 sets of previously published critiques of the quantum approach (in particular the Penrose-Hameroff Orch OR model), and reply accordingly. But first, let's examine some of the assumptions implicit in the critiques. For example the question is often posed: how could quantum states persist in the "warm, wet, noisy environment within the brain?"

3. Assumptions: Is the brain "warm, wet and noisy?"

a. Is the environment warm?

Yes, clearly the brain operates at 37.6 degrees centigrade, and deviations in brain temperature in either direction are not well tolerated for consciousness. This temperature is quite toasty compared to the extreme cold needed for quantum technological devices which operate near absolute zero, or at best sub-Siberian temperatures. In technology, the extreme cold serves to prevent thermal excitations which could disrupt the quantum state. However proposals for biological quantum states suggest that biological heat is funneled (Conrad) or used to pump coherent

excitations (Frohlich, or Fermi-Pasta-Ulam resonance, or Davydov solitons). In other words biomolecular systems may have evolved to utilize thermal energy to drive coherence.

A related question is whether consciousness is dissipative. Clearly biological processes are dissipative in general, but there's some suggestion that consciousness is not dissipative.

Evidence from brain imaging shows a discrepancy between increased blood flow and oxygen uptake in areas actively involved in cognitive processing (and presumably neural correlates of consciousness). Large increases in blood flow are accompanied by little or no increase in oxygen uptake. This has prompted some brain imagers (e.g. Peter Fox at San Antonio) to suggest that consciousness is anaerobic---does not require oxygen. As everyone knows, the brain cannot function or survive without oxygen for more than seconds or minutes, and so at first glance the suggestion seems preposterous. However as Fox points out, the energy cost of a phone system is largely in setting up and maintaining lines of communication; communication itself (talking on the phone) is very cheap energetically. Aerobic processes may be necessary to set the stage for consciousness which, in and of itself may be anaerobic and non-dissipative. (Quantum computation must be non-dissipative). How could this be?

One possibility is that processes related to consciousness are phasic, with cycles of aerobic, dissipative classical processes alternating with 2) anaerobic, non-dissipative quantum processes (consciousness is the transition from 2 to 1). Phase 2 would be the pre-conscious quantum phase isolated from environment by actin gelation (see later section). This is relevant because quantum computation must be reversible, and non-dissipative. Quantum computation would thus be well suited to an anaerobic, non-dissipative phase. Input/output environmental interactions would be well suited to an aerobic, dissipative classical phase.

Figure 1. Schematic sequence of phases of actin gelation/quantum isolation (13) alternating with phases of solution/environmental communication (4) surrounding microtubules. Cycles may occur rapidly, e.g., 25 msec intervals (40Hz).

Figure 2. An Orch OR event. a) Microtubule simulation in which classical computing (step 1) leads to emergence of quantum coherent superposition (and quantum computing (steps 23) in certain (gray) tubulins. Step 3 (in coherence with other microtubule tublins) meets critical threshold related to quantum gravity for selfcollapse (Orch OR). A conscious event (Orch OR) occurs in the step 3 to 4 transition. Tubulin states in step 4 are noncomputably chosen in the collapse, and evolve by classical computing to regulate neural function. b) Schematic graph of

proposed quantum coherence (number of tubulins) emerging versus time in microtubules. Area under curve connects superposed mass energy E with collapse time T in accordance with $E=h/T$. E may be expressed as Nt , the number of tubulins whose mass separation (and separation of underlying space time) for time T will selfcollapse. For $T = 25 \text{ msec}$ (e.g. 40 Hz oscillations), $Nt = 2 \times 10^{10}$ tubulins.

In summary, the brain is indeed warm. The assumption/prediction by quantum advocates is that biological systems (at least those with crystal lattice structures) have evolved techniques to funnel thermal energy to coherent vibrations conducive to quantum coherence, and/or to insulate quantum states through gelation or plasma phase screens.

b. Is the environment wet?

The conventional picture of living cytoplasm (cell interior) is a teeming minestrone of biomolecules. However cytoplasm actually exists in different phases of 1) liquid "sol" (solution, minestrone soup), and solid "gel" (gelatinous phases of various sorts, "jello"). Transition between sol and gel phases depends on actin polymerization. Triggered by changes in calcium ion concentration (or photons in primitive cells), actin co-polymerizes with different types of "actin cross-linking proteins" to form dense meshworks of microfilaments and various types of gels which encompass microtubules and organelles (soup to jello). Characteristics of the actin gels are determined by the particular type of actin cross-linkers. Gels depolymerize back to liquid phase by calcium ions activating gelsolin protein which severs actin (jello to soup). Actin repolymerizes into gel when calcium ion concentration is reduced (soup to jello). Actin gel, ordered water "jello" phases alternate with phases of liquid, disordered soup. Exchange of calcium ions between actin and microtubules (and microtubule-bound calmodulin) can mediate such cycles.

Figure 3 Actin monomers self-assemble suddenly into filaments which form a meshwork. Lower left shows actin gel mesh encompassing hidden microtubules.

The sol-gel transition is a very primitive biological phenomenon, related to movement of single cells like amoeba or our own lymphocytes. In these cases the actin polymerizes in one direction, and liquefies behind it, causing a directional flowing of cytoplasm. (Actin also polymerizes in response to light). In the 19th century Claude Bernard studied this phenomenon, which he called cytoplasmic streaming, and discovered it was sensitively inhibited by exposure to the anesthetic gas chloroform. In more complex, asymmetrical cells like neurons, transport and motion utilize polymerization of microtubules, though actin gelation still plays a role. Even in the "sol", or liquid phase, water within cells is not truly liquid and random. Pioneering work by Clegg (1984) and others have shown that water within cells is to a large extent "ordered," and plays the role of an active component rather than inert background solvent. Neutron diffraction studies indicate several layers of ordered water on such surfaces, with several additional layers of partially ordered water.

Protein-water binding/ordering is well studied, and linked to events in hydrophobic pockets on the protein interior. Wulf and Featherstone (1967) showed that anesthetic binding in a hydrophobic pocket altered the number of water molecules bound at the protein surface.

Water molecules bound to actin and other cytoskeletal surfaces, should also be ordered and coupled to the actin/cytoskeletal dynamics. (See Scott Hagan's contribution for a description of quantum field theoretical approaches to the ordered states of biological water). Watterson (1981, 1996) observes that in the conventional view of gelation, long cross-linked polymer solutes form a spacious network. But in living cytoplasmic gels, the water doesn't flow - even though the gel is over 75% water. NMR studies have shown that actin assembly results in reduced water mobility (ordering), and that distribution of ordered water through the cell is a heterogeneous and dynamic process. Pauser et al (1995) demonstrated that 55% of the water of the vegetal pole region of frog oocytes is bound water, with less bound (~25%) near the animal pole cytoplasm, and ~10% bound in the nucleus. Ordered water distribution changes in time also. For example cell cycle (mitosis) changes correlate with actin polymerization, gelation, and reduced cytoplasmic water motion (Cameron et al, 1987).

The character of actin gelation and water ordering depends on actin cross-linking. Of the various cross-linker related types of gels, some are viscoelastic, but others (e.g. those induced by the actin cross-linker avidin) are solid and can be deformed by an applied force without any response (Wacchstock et al, 1994). Such shock-absorbance would be useful in quantum isolation.

Cycles of actin gelation/solution can be quite rapid, occurring for example at 40 Hz. In neurons Miyamoto (1995) and Muallem et al (1995) have shown that cycles of actin gelation/solution correlate with release of neurotransmitter vesicles from pre-synaptic axon terminals. I'm unaware of any studies of sol-gel transitions in dendrites (a good literature review research question for any interested students!).

So the environment is not necessarily wet, at least it's not wet some of the time.

c. Is the environment noisy?

Let's consider two types of noise: a) intracellular noise, presumably from thermal energy of water, and b) electrical noise as manifest in electrophysiological recordings.

Several proposals have been put forth suggesting that thermal energy in biological systems is somehow transformed into coherent excitations in biomolecules. Michael Conrad discussed "funneling" of thermal energy, and the Davydov soliton has been suggested and discussed for many years. Perhaps the best known suggestion is that of Frohlich.

Herbert Frohlich, an early contributor to the understanding of superconductivity, predicted quantum coherence in living cells (based on earlier work by Oliver Penrose and Lars Onsager (1956) Frohlich (1968, 1970, 1975) theorized that sets of protein dipoles in a common electromagnetic field (e.g. proteins within a polarized membrane, subunits within an electret polymer like microtubules) undergo coherent conformational excitations if thermal energy is supplied. Frohlich postulated that biochemical and thermal energy from the surrounding "heat bath" provides such energy. Cooperative, organized processes leading to coherent excitations emerged,

according to Frohlich, because of structural coherence of hydrophobic dipoles in a common voltage gradient.

Coherent excitation frequencies on the order of 10⁹ to 10¹¹ Hz (identical to the time domain for functional protein conformational changes, and in the microwave or gigaHz spectral region) were deduced by Frohlich who termed them acousto-conformational transitions, or coherent (pumped) phonons, or optical phonons. Such coherent states are termed Bose-Einstein condensates in quantum physics and have been suggested by Marshall (1989) to provide macroscopic quantum states which support the unitary binding of consciousness.

As Al Scott described, experimental searches for Frohlich coherence have been confusing, and to some extent disappointing. However biological mechanisms developed to isolate and protect quantum coherence mechanisms could also make their detection quite difficult. Nevertheless experimental evidence for Frohlich-like coherent excitations in biological systems includes observation of gigaHz-range phonons in proteins (Genberg et al, 1991) sharp-resonant non-thermal effects of microwave irradiation on living cells (Grundler and Keilman, 1983; Genzel et al, 1983), gigaHz induced activation of microtubule pinocytosis in rat brain (Neubauer et al, 1990), and laser Raman spectroscopy detection of Frohlich frequency energy (Vos et al, 1993).

Also, see Scott Hagan's contribution for a description of how quantum field theory could order intracellular cytoplasm.

The second type of noise commonly assumed is electrical. Electrophysiological recordings of single cell events detect fluctuations in baseline voltage over various time scales. To get good signals from individual neurons, multiple recordings are done over time and the background fluctuations averaged out. The assumption is that the background fluctuation is meaningless noise. However Aviram Grinvald and his group in Israel have measured this background (Grinvald terms it ongoing activity) over wide regions of brain and found the activity to be correlated. It only appears random locally! Over the whole brain the activity is coupled, and thus may not be noise at all.

One other possible factor would be the presence of plasma-like charge layers in cytoplasm. Triffet and Green have suggested that double layers of opposite charges along membrane and other surfaces form a plasma-like "Debye layer" zone with perhaps superconductive or other quantum behaviors. A similar effect may occur on the surface of microtubules. Dan Sackett at NIH has shown that at precisely physiological pH the terminal chain of amino acids on each tubulin within microtubules sticks out into cytoplasm. Each "hair"-like projection carries 8 negatively charged amino acids, and apparently attracts 8 positive charges on ions (e.g. 4 calcium ions). The counterions are expected to dance cooperatively in a quantum state plasma phase which, according to Sackett, could screen quantum states in microtubules from environmental decoherence outside the plasma layer, or host relevant quantum mechanisms in the MT surface plasma itself. (Sackett points out that more interesting behavior occurs on the surface of the earth than within it).

4 Negatively charged terminal amino acids extend from each tubulin into cytoplasm, attracting counter-ions, forming a quantum state plasma-like charge layer according to mechanism proposed by Dan Sacket at NIH.

Sacket's plasma sleeve may solve an old mystery in microtubule history. In many electron micrographs of microtubule cross section the MT is surrounded by a "clear zone" of several nanometers which excluded organelles, the staining agent and apparently everything else observable. Yet in many other micrographs no such clear zone was present around microtubules. If Sacket is correct the plasma layer causes the clear zone, but only under very specific conditions of pH.

So the "warm, wet and noisy" brain may not be as hostile an environment for ordered quantum states as it is reputed to be.

4. Summary of the Orch OR model

Published criticisms of quantum effects relevant to consciousness will be considered in Section 5. As several of these are aimed specifically at the Penrose-Hameroff Orch OR model, the key points of Orch OR are presented here. The full story is available at [~http://www.u.arizona.edu/~hameroff/royal.html](http://www.u.arizona.edu/~hameroff/royal.html) [<http://www.u.arizona.edu/hameroff/royal.html>](http://www.u.arizona.edu/hameroff/royal.html) and elsewhere.

Conformational states of individual tubulin proteins in brain microtubules are sensitive to internal quantum events (e.g. London forces in hydrophobic pockets) and able to cooperatively interact with other tubulins in both classical and quantum computation (Figures 5-9). Classical phase computation (microtubule automata) regulates chemical synapses and other neural membrane activities.

Figure 5. Left: Microtubule (MT) structure: a hollow tube of 25 nanometers diameter, consisting of 13 columns of tubulin dimers arranged in a skewed hexagonal lattice (Penrose, 1994). Right (top): Each tubulin molecule may switch between two (or more) conformations, coupled to London forces in a hydrophobic pocket. Right (bottom): Each tubulin can also exist in quantum superposition of both conformational states (Figure 1a, c.f. Hameroff and Penrose, 1996).

Figure 6. Microtubule automaton simulation (from Rasmussen et al., 1990). Black and white tubulins correspond to black and white states shown in Figures 1a and 3. Eight nanosecond time steps of a segment of one microtubule are shown in

"classical computing" mode in which conformational states of tubulins are determined by dipole-dipole coupling between each tubulin and its six (asymmetrical) lattice neighbors calculated by where y_i and r_i are intertubulin distances, e is the electron charge, and ϵ is the average protein permittivity. Conformational states form patterns which move, evolve, interact and lead to emergence of new patterns.

Figure 7. Schematic of quantum computation of three tubulins which begin (left) in initial classical states, then enter isolated quantum superposition in which all possible states coexist. After reduction, one particular classical outcome state is chosen (right).

Figure 8. Microtubule automaton sequence simulation in which classical computing (step 1) leads to emergence of quantum coherent superposition (steps 2-6) in certain (gray) tubulins due to pattern resonance. Step 6 (in coherence with other microtubule tubulins) meets critical threshold related to quantum gravity for selfcollapse (Orch OR). Consciousness (Orch OR) occurs in the step 6 to 7 transition. Step 7 represents the eigenstate of mass distribution of the collapse which evolves by classical computing automata to regulate neural function.

Quantum coherence begins to reemerge in step 8.

Quantum coherent superposition supporting quantum computation emerges among London forces in hydrophobic pockets of microtubule subunit tubulins (e.g. in a manner described by Frohlich, 1968; 1975). In this phase, quantum computation among tubulins evolves linearly according to the Schrodinger equation (quantum microtubule automata). Actin gelation and a condensed charge phase surrounds, isolates and insulates microtubules from thermal/environmental decoherence during the quantum phase.

The proposed quantum superposition/computation phase in neural microtubules corresponds to preconscious (implicit) processing, which continues until the threshold for Penrose's objective reduction is reached. Objective reduction (OR)-a discrete event then occurs (Figures 5-7), and post - OR tubulin states (chosen non-computably) proceed by classical microtubule automata to regulate synapses and other neural membrane activities. The events are proposed to be conscious (to have qualia, experience) for reasons that relate to a merger of modern physics and philosophical pan-experientialism (see earlier lectures). A sequence of such events gives rise to a stream of consciousness.

Microtubule quantum states link to those in other neurons and glia by tunneling through gap junctions (or quantum coherent photons traversing membranes - Jibu

and Yasue, 1995; Jibu et al, 1994; 1996). This spread enables macroscopic quantum states in networks of gap junction connected cells (neurons and glia) throughout large brain volumes (Figure 9).

9. Schematic of proposed quantum coherence in microtubules in three dendrites interconnected by tunneling through gap junctions. Within each neuronal dendrite, microtubule associated protein (MAP) attachments breach isolation and prevent quantum coherence; MAP attachment sites thus act as "nodes" which tune and orchestrate quantum oscillations and set possibilities and probabilities for collapse outcomes (orchestrated objective reduction: Orch OR). Gap junctions may enable quantum tunneling among dendrites in macroscopic quantum states. Dendritic lamellar bodies may act as tunneling diodes on opposite sides of gap junctions in neurons (see previous week's lecture).

Probabilities and possibilities for pre-conscious quantum superpositions are influenced by biological feedback including attachments of microtubule associated proteins ("MAPs"), which tune and "orchestrate" quantum oscillations (Figure 9). We thus term the selftuning OR process in microtubules "orchestrated" objective reduction Orch OR.

Orch OR events may be of variable intensity and duration of preconscious processing. Calculating from $E = h/T$, for a preconscious processing time of e.g. $T = 25$ msec (thalamo-cortical 40 Hz), E is roughly the superposition/separation of 2×10^{10} tubulins. For $T = 100$ msec (alpha EEG) E would involve 5×10^9 tubulins. For $T = 500$ msec (e.g. shown by Libet et al., 1979, as a typical preconscious processing time for low intensity stimuli), E is equivalent to 10^9 tubulins. Thus 2×10^{10} tubulins maintained in isolated quantum coherent superposition for 25 msec (or 5×10^9 tubulins for 100 msec, or 10^9 tubulins for 500 msec, etc.) will selfcollapse (Orch OR) and elicit a conscious event.

Each brain neuron is estimated to contain about 10^7 tubulins (Yu and Bass, 1994). If, say, 10 percent of each neuron's tubulins became coherent, then Orch OR of tubulins within roughly 20,000 (gap junction connected) neurons would be required for a 25 msec conscious event, 5,000 neurons for a 100 msec event, or 1,000 neurons for a 500 msec event, etc.

Each instantaneous Orch OR event binds superposed information encoded in microtubules whose net displacement reaches threshold at a particular moment: a variety of different modes of information is thus bound into a "now" event. As quantum state reductions are irreversible in time, cascades of Orch OR events present a forward flow of time and "stream of consciousness."

The key points relevant to the present discussion are that Orch OR predicts that quantum states in microtubules are isolated from environmental decoherence for 25 msec and longer due to gelation surrounding MTs, plasma phase screening, ordered water, and Frohlich coherence. Spread among neurons is predicted to occur by tunneling through

gap junctions, though the Jibu/Yasue/Hagan predicted coherent domain of 50 microns would also work.

5. Feasibility of quantum states in the brain:

reply to quantum critics Tegmark, Tuszynski and Scott

A number of published articles have criticized the concept of quantum effects relevant to consciousness, and in some cases elicited published replies. For entertainment I'd recommend the article "Gaps in Penrose's toilings" by Rick Grush and Patricia Churchland in *Journal of Consciousness Studies*, and the reply article "What gaps? Reply to Grush and Churchland" by Roger Penrose and me in the subsequent issue of JCS. At the close of their article (despite admittedly having not read any details of the Penrose-Hameroff Orch OR model) Grush and Churchland conclude: "the Penrose-Hameroff model is no better supported than one in a gazillion caterpillar-with-hookah hypotheses". Taking the Alice in Wonderland bait we reply at the end of our article "it's not that we're in wonderland, but p'raps their heads are in the sand". The *Journal of Consciousness Studies* then commissioned and published as an accompaniment to our article a cartoon which they felt captured the (gutter-level) spirit of the debate. See Figure 10.

Figure 10.

For a continuation of the intellectual food fight I'd also recommend Pat Churchland's chapter in the Tucson II book : "On non-neural theories of the mind" and my reply piece "More neural than thou". However none of these directly pertain to the issue at hand - whether quantum states, and in particular macroscopic quantum states, are bioenergetically feasible in the brain. To best address this question I consider here publications by 3 worthy adversaries and knowledgeable critics, two of whom (Jack Tuszynski and Alwyn Scott) are obviously commentators here. The first, however, is Max Tegmark

Max Tegmark

In an attempt to refute quantum models of consciousness, physicist Max Tegmark at the Institute for Advanced Studies in Princeton has written a paper "The quantum brain" which is posted on the Los Alamos Archives under quant-ph/9907009 5 July 1999.

Tegmark's main point may be summarized:

"The make or break issue for all these quantum models is whether the relevant degrees of freedom of the brain can be sufficiently isolated to retain their quantum coherence, and opinions are divided. For instance Stapp has argued that interaction with the environment is probably small enough to be unimportant for neural processes whereas Hawking and Scott have conjectured that environment-induced coherence will rapidly destroy macrosuperpositions in the brain. It is therefore timely to try and settle the issue with detailed calculations of the relevant decoherence rates. This is the purpose of the present work."

But what are the relevant degrees of freedom?

Tegmark gives two treatments of quantum proposals: 1) superpositions of neurons firing and not firing, 2) superpositions of the location of a soliton on a microtubule. He

calculates decoherence times due to interaction with environmental ions as 10-20 sec for superpositions of neurons firing/not firing, and 10-13 seconds for superpositions of solitons on microtubules.

I agree with his assessment that superpositions of neurons firing and not firing is unlikely. However regarding microtubules, Tegmark considers a model of classical kinks/solitons traveling along microtubules published by Sataric et al (based on our classical microtubule automaton model). Though he targets Penrose, Tegmark ignores the specifics of the Penrose Hameroff Orch OR proposal.

Tegmark considers a kink/soliton in superposition of two different locations along the microtubule and then calculates interactions between the soliton displacement and calcium ions associated with the microtubule. If Tegmark claims to be criticizing Penrose's view, then he should at least be familiar with the details of the Orch OR proposal which he doesn't even mention or reference. So the degrees of freedom he is using in "refuting Penrose" is from a quantum model he himself has invented

Interestingly, Tegmark ignores interactions between proposed quantum states and surrounding water, taking the suggestion that surrounding water is "ordered" (though he attributes this notion to Nanopoulos and Mavromatos, who got the idea from Penrose-Hameroff).

As far as ions inducing decoherence, this is a bit puzzling if Tegmark is assuming the water is ordered. Because if the water is ordered then the ions in the water (depending on their size relative to water) are also ordered. Ions whose radius is smaller than the H₂O radius (1.38 angstroms) do not disturb the ordering (Ergin, 1983; Uedaira and Osaka, 1989; Jibu et al, 1995). Sodium ions (radius 0.98 angstroms), calcium ions (1.0 angstroms) and magnesium ions (0.72 angstroms) can all embed in ordered water without disturbance. Ions whose radius is close to that of water (e.g. potassium 1.38 angstroms) can replace water molecules without disturbance, whereas larger ions will disturb ordering. Chloride (1.81 angstroms) is in the latter category and should disrupt water ordering. Chloride intra-cellular concentration is extremely low except for terminal phase of an action potential.

In any case Tegmark calculates a decoherence time for his superpositioned kink/solitons of 10-13 seconds based on an equation with the following characteristics (eqs 19 and 22 in his paper). The decoherence lifetime is related to an expression which has in the numerator the distance a^3 . a is the distance from the superposition to the nearest ion. Since ions may be embedded in ordered water, or separated from the superposition by a gel state, a in the Orch OR context at least may be some distance away. As a cubed is in the numerator, a 100 fold increase in effective a (e.g. from 0.1 nanometers to 10 nanometers) would be 6 orders of magnitude lengthening of the coherence time.

The denominator in Tegmark's equation includes the separation distance between the separated superpositioned kink/solitons which he takes to be several tubulin lengths, or roughly 24 nanometers. However in the Orch OR model the separation of tubulin proteins from themselves only requires the diameter of one atomic nucleus, or fermi length (~10⁻⁶ nanometers). This is 7 orders of magnitude smaller in the denominator, combined with 6 orders of magnitude larger in the numerator. Compared to Tegmark's 10-13 seconds, this gives us on the order of roughly 1 second. Orch OR requires 25 msec or longer. I'd be happy to hear from physicists and others who would look at Tegmark's paper and are familiar with Orch OR.

Figure 11 In the Orch OR model (see ~<http://www.u.arizona.edu/~hameroff/or.html> <<http://www.u.arizona.edu/hameroff/or.html>>) we calculated gravitational self-energy for superposition/separation of tubulin in 3 ways: as protein spheres, at the level of atomic nuclei, and at the level of individual nucleons (protons and neutrons). E was highest (thus quickest, dominant collapse) for separation at the level of atomic nuclei.

b. Tuszynski and Brown

A special issue of the Philosophical Transactions of the Royal Society is devoted to quantum computation. It includes a pair of articles about the possibility of quantum computation in microtubules---one positive (by me), and one skeptical by Jack Tuszynski and Andrew Brown. Jack and Andrew raise several specific points which I reply to in an appendix to my paper (on my website URL. Here are Jack and Andrew's criticisms (*in italics*) followed by my responses.

Gravitational effects should be entirely overshadowed by the remaining processes.

The energy from an Orch OR event is indeed very small compared to thermal noise (kT) and would seemingly drown in an aqueous medium. Isolation/insulation mechanisms are thus required to shield microtubules from thermal noise or any type of environmental decoherence. The Orch OR model suggests that quantum coherent superposition occurs in microtubules which are immediately surrounded by an insulating charge condensation and encased (cyclically) in actin gelation (Section III). Cyclical isolation allows for alternating phases of communication (input/output) and isolated quantum computation. In addition to isolation, microtubule subunits (tubulins) must also be sensitive to quantum influences from other superposed tubulins and non-computable influences in Planck scale geometry. In questioning the robustness of proposed quantum effects, Tuszynski and Brown ascribe the gravitational energy for a tubulin protein in Orch OR to be the attraction between two masses given by the standard Gm^2/r , where G is the gravitational constant, m is the mass of tubulin, and r is the distance between the two masses which Tuszynski and Brown take to be the radius of tubulin. This would accurately describe the gravitational attraction between two adjacent tubulins (or tubulin monomers), and yields an appropriately small energy of 10^{-27} eV. However the relevant energy in Orch OR is the gravitational self-energy E of a superposed mass m separated from itself by distance a , given (for complete separation) by $E=Gm^2/a$. In Hameroff and Penrose (1996) we calculated this energy for three cases: 1) partial separation of the entire protein by one tenth its radius, and 2) complete separation at the level of each proteins atomic nuclei ($a=2.5$ fermi lengths), and 3) complete separation at the level of each protein's nucleons ($a=0.5$ fermi). Of these, highest energies were for separation at the level of atomic nuclei, roughly 10^{-21} eV per tubulin (although separation at the level of, say, atoms or amino acids may yield higher energy). As roughly 2×10^{10} tubulins are involved in each proposed Orch OR event (e.g. for superpositions lasting 25 msec) the energy is on the order of roughly 10^{-10} eV, or 10^{-28} joules, still extremely tiny (kT is about 10^{-4} eV). However the 10^{-28} joule energy emerges abruptly, e.g. within one Planck time of 10^{-43}

seconds. This may be equivalent to an instantaneous jab of 1013 watts (joules/sec), roughly 1 kilowatt per tubulin per conscious event.

The size of the tubulin protein is probably too large to make quantum effects easily sustainable.

Nanometer size proteins such as tubulin (8 nm x 4 nm x 4 nm) may be optimal scale for a quantum/macrosopic interface (Watterson 1991, Conrad 1994). Smaller biomolecules lack causal efficacy of structural protein conformational changes responsible for a host of biological functions. Larger molecules would be insufficiently sensitive to quantum effects.

Conformational effects are expected to involve distances of 10 angstroms (1 nanometer), larger than those called for in the Orch OR model.

The superposition separation distance (e.g. 1 atomic nucleus, 10⁻⁶ nanometer in the case cited) is indeed much smaller than conformational changes which may approach 1 nanometer. As described in Section II proteins are relatively unstable and their conformation regulated through nonlinear "quakes" mediated through quantum-level London forces.

Physiological temperature requirements make it extremely difficult to defend the use of the quantum regime due to the persistence of thermal noise.

A biological quantum state must be isolated/insulated from thermal noise, or funnel it into coherence, features nature may have evolved in cytoplasmic actin gelation and condensed charged layers (Section III). Some evidence supports biological quantum states (e.g. Tejada et al, 1996; Waliczek, 1995). According to the Frohlich mechanism, thermal energy in biological systems may condense to a coherent mode.

...microtubules are extremely sensitive to their environment...we doubt that microtubules can be shielded.

As described in Section III, nature may have solved the problem of both isolation and communication by alternating cytoplasmic phases of solution ("sol", liquid, sensitive to environment, classical) and gelation ("gel", solid, shielded/insulated, quantum). Thus microtubules can be both sensitive to their environment ("sol" phase) and isolated/shielded ("gel" phase).

...two (or possibly more) conformational states of tubulin are separated by a sizable potential barrier which again requires an external stimulus (such as GTP hydrolysis) to overcome it.

Tubulin has numerous possible conformations which can interchange without GTP hydrolysis (Section II). The two state tubulin model is a simplification. The structure of tubulin has recently been clarified (Nogales et al, 1998) so molecular simulations will soon be available.

...the 500 msec preconscious processing time may be directly related to the action potential travel time along the axon plus the refractory lag time in synaptic transmission rather than to the quantum collapse time.

In the Orch OR model the "quantum collapse time" T is chosen to match known neurophysiological time intervals related to pre-conscious processes; the gravitational self-energy E and related mass may then be calculated. For example we have used 25 msec (e.g. in coherent 40 Hz oscillations), 100 msec (e.g. EEG alpha rhythm), and 500 msec (e.g. Libet's pre-conscious threshold for low intensity sensory stimuli).

If quantum superposition correlates with pre-conscious processing, then dendritic activities (more than axonal firings) are likely to be relevant to consciousness (e.g. Pribram, 1991). Microtubules in dendrites are of mixed polarity (unlike those in axons), an arrangement conducive to cooperative computation.

Tuszynski and Brown raise valid objections; quantum states in a biological milieu appear at first glance to be unlikely. However nature may have evolved specific conditions for isolation, thermal screening and amplification. Life itself may be a macroscopic quantum state.

The third critique is Al Scott's article "On quantum theories of the mind" in JCS 1996, 3(5-6) 484-491. Al's main points (*italics*) are the following.

Brain activities relevant to consciousness are nonlinear, and quantum theory is linear (implying classical nonlinear dynamics is more fertile ground).

While the Schrodinger equation and the evolution of the quantum wave function may be linear, the collapse of the wave function---particularly in the Penrose 'objective reduction' formulation--- is decidedly nonlinear.

Born-Oppenheimer method - Al uses the Born-Oppenheimer method to conclude that quantum effects are insignificant in molecular dynamics.

The Born-Oppenheimer approximation concerns the relative influences of an atomic nucleus and its surrounding electrons, and treats them something like the earth and a soccer ball, respectively. That is, the nucleus may be considered stationary, able to influence the electron whereas the electron moves but is unable to influence the nucleus. Conrad points out that although the mass of nucleus is far greater than that of the electrons, net charge is equal. Electrons which delocalize---that is travel among resonance orbitals of several atoms such as an aromatic ring in a hydrophobic amino acid such as tryptophan---can indeed influence nuclei, and hence protein conformation.

Quantum mechanical wave length. Al calculates that due to the mass of tubulin, quantum delocalization would only extend a fraction of an atomic diameter.

In Orch OR we claim superposition---separation---of each tubulin by one atomic radius, roughly 1/4000 of an atomic diameter. So a fraction of an atomic diameter is OK.

Hodgkin-Huxley equations. Al describes how the purely classical H-H equations accurately describe the nonlinear propagation of an action potential along an axon without need to resort to quantum effects.

No problemo. Axonal action potentials may not be directly involved in consciousness (though cytoskeletal activities in shadowy concert with axonal membrane activities might be). There are many brain activities which have nothing to do with consciousness.

Schrodinger's cat. Al describes how Schrodinger devised his famous thought experiment to illustrate the inappropriateness of applying quantum effects to biology.

We know the story---a cat is in a box which has a poison vial. A microscopic quantum event, e.g. passage of a photon through a half silvered mirror is coupled to the poison. According to (the Copenhagen interpretation of) quantum theory, the photon both passes through, and does not pass through the mirror (and both triggers and doesn't trigger the poison). Therefore until the box is opened and consciously observed the cat is both dead and alive.

Despite the ridiculousness of the scenario, the answer is still somewhat puzzling. Al offers his objections, one of which I'll comment upon. He states that "a conservative estimate suggests that a time very much longer than the age of the universe would be

required for the cat's wave function to rotate from being dead to being alive. For all practical purposes this implies that a quantum mechanical cat would be either dead or alive".

On the contrary, in Penrose objective reduction an isolated superposition will self-collapse by $E = h/T$, where T is the time until collapse, h is Planck's constant over 2π , and E is the degree of superposition. Small masses in superposition will self-collapse only after a very long time (for an isolated superposed electron, 10 million years). An isolated superposed cat of roughly 1 kilogram would self-collapse in 10^{-37} seconds. For all practical purposes this implies that a quantum mechanical cat would be either dead or alive.

6. Conclusion

Quantum approaches have a great deal of explanatory power for consciousness. Although at first glance the possibilities of macroscopic quantum states in the brain seem slim, there is reason to believe that nature has evolved specific mechanisms to isolate and support such states. The proposals described in Orch OR and other models are testable, if not presently then in the foreseeable future. Attempts to dismiss quantum proposals out-of-hand or by incorrect assumptions fail. Alternative explanations based on classical emergence make no testable predictions. Time will tell.

Quantum Field Theoretical Approaches to Consciousness

Scott Hagan

Introduction and Motivation

Rather than attempting to find a role for consciousness in the interpretation of quantum mechanics, quantum field theoretical approaches instead use the standard machinery of contemporary quantum theory to attempt to understand, in particular, the unified wholeness characteristic of consciousness (James, 1890/1918) but nowhere in evidence in the classical world. From their earliest roots (Ricciardi and Umezawa, 1967), the picture of consciousness that has arisen in these quantum models has been intimately linked to memory. It was an understanding of its enigmatic features that prompted the first model (Stuart *et al.*, 1978, 1979) in which memory was envisioned as a macroscopic ordered state. The suggestion was that the brain instantiates not one, but two mechanisms of memory, that it constitutes a mixed quantum-classical system. Aspects of memory that elude understanding from a classical view might be naturally explained in a quantum picture and *vice versa*. The view that a second mechanism might be operative in the brain, responsible for non-local (in the neurophysiological, not the physical sense) properties of memory, has long been advocated by Pribram (1971, 1991), who suggested dendritic nets as the locus for an encoding in terms of interference patterns analogous (not literally identical) to the storage mechanism in holograms.

In the quantum field theoretical approaches to be discussed, consciousness is associated to specific imprinting and recall mechanisms. Under the action of an external stimulus, the excitation of certain long-range correlation modes (called *symmetron* modes in the work of Stuart *et al.*) corresponds to the brain "consciously feeling" a complex pattern encoded in the vacuum (ground state) structure. The conscious moment becomes a sort of "remembered present" (Edelman, 1990). In contemporary models, memories are imprinted with widely varying decay times so that it may be necessary to refresh memories that decay more rapidly, lest they be forgotten (Vitiello, 1995) and some memories may be more difficult to excite than others (Hagan and Hirafuji, 1998).

It was clear from the outset that neurons would not be the appropriate functional units in any realization of macroscopic ordered states in the brain. A picture of how theories of this kind might be physically instantiated in the brain has arisen from several quarters, following early suggestions by Frohlich (1968) that condensation in biological media can be mediated by a metabolic supply of energy above a threshold value. Before discussing these, it will be helpful to digest some relevant concepts.

Spontaneous Symmetry Breaking

All of these theories take the notion of *spontaneous symmetry breaking* as their founding principle. This is an idea, familiar in both high energy and condensed matter physics, that allows a rich vacuum structure to be instantiated. In quantum mechanics, there is always only one vacuum, a unique lowest energy state, a result known as the Equivalence Theorem. In field theory this need not be the case, and this fact allows inhomogeneous media to be encompassed by the theory: the dynamics in different regions of space may "act out of" different vacua, and therefore describe radically different behavior in the same medium.

Spontaneous symmetry breaking occurs whenever some certain symmetry of the dynamics of a system, some transformation of the variables that leaves the equations invariant, is not also a symmetry of the vacuum. To visualize an example of this, imagine

an "energy landscape" associated with every point in space. If the landscape were to have the shape of a bowl, then the dynamics would be invariant under two-dimensional rotations about the axis of symmetry. The vacuum, the unique lowest energy state at the bottom of the bowl, would likewise be invariant under such rotations. But imagine instead that the energy landscape has a dimple in it (see Figure 1). Here the dynamics are likewise invariant under rotations, but now the vacuum state might be any one of an infinite number of states lying along the bottom of a circular trough. Whichever one is chosen will no longer be rotationally invariant, a spontaneous symmetry breaking scenario. We might even imagine this dimpled landscape with another dimple (in the opposite direction) at its center. This would then have one ordinary, rotationally invariant vacuum and an infinite number of spontaneous symmetry breaking vacua. The symmetry generally invoked in current models is the three-dimensional rotational symmetry of electric dipoles (actually a related but not quite equivalent symmetry of molecular dipole fields). Though water, the most ubiquitous polar molecule in biological media, is an obvious candidate, many biomolecules are polar and might also be considered in this role.

Figure 1: An example of spontaneous symmetry breaking. Both massive and massless modes are generically present.

In the case of a spontaneously broken symmetry, the so-called "normal" modes comprise both massive and massless modes. A massive mode describes the possible effect on the system when it is excited by an input of energy. In our dimpled example, these are modes that take the system point up the walls of the trough. The massless modes describe the movement of the system point along the trough and can be excited with no energy input since every point in the trough is equally a state of lowest energy. These massless modes, known as *Nambu-Goldstone bosons*, are the symmetron modes of Stuart *et al.* (1978, 1979) and are components of the system only in a rather odd sense. They would not be listed in any inventory of parts if one were to disassemble the system but rather appear in the system only dynamically. Their existence allows the system to effect a sort of restoration of the spontaneously broken symmetry. Through the exchange of these bosons, it is ensured that a coherent choice for the vacuum is made throughout the scope of the system (long-range correlation) and because the modes are massless, this requires no energy. Each vacuum choice is associated with a macroscopic phase.

In the case of a closed system, the enforcement of a particular choice for the vacuum is in fact so rigid that it is impossible to move between them (they are unitarily inequivalent). In the model of memory encoding proposed by Stuart, Takahashi and Umezawa, (1978, 1979) this would suggest, for instance, that it should be impossible to forget. While this is clearly not the case, the brain is clearly not a closed system. In field theory, a closed system is, in effect, treated as one of infinite extent – the boundary between system and environment is moved out to infinity so there is no possibility of interaction. The realistic case of an open, dissipative system (Vitiello, 1995) amends this to a finite volume. In the process one loses some of the rigid stability of the vacuum in the infinite volume case, so

it becomes possible to move between vacua and hence, forget. The Nambu-Goldstone bosons also acquire a very small mass, generally indicated by referring to them as "pseudo-Goldstone bosons". The smallness of this mass allows very weak perturbations to play a role in effecting phase transitions between vacua and .

Condensation

According to a quantum field theoretical model, memory is established and coded by the condensation of these modes in particular vacua. *Condensation* is a phenomenon that results from the "sociable" character of bosons. Unlike "antisocial" fermions, which never deign to appear in the same state as any other fermion, bosons will, under particular conditions, cluster in the ground state, acting as a unified whole. When and where this behavior will occur is controlled by a balance of several parameters, most notably temperature, condensate density and the mass of the condensing species (other factors, like pressure, also come into play to a lesser degree).

In the most famous example of condensation, superconductivity, the operative species is not the electron, which is a fermion and hence loath to participate. At very low temperatures, however, electrons can overcome their mutual repulsion sufficiently to form pairs (Cooper pairs) that then act like bosons, and it is these pairs that condense. The temperature below which this will occur is inversely proportional to the mass of the Cooper pair which, in condensation terms, is relatively high so that the temperature is generally extremely low. It is not currently well-understood why the so-called high-Tc superconductors can undergo this transition at much higher temperatures (still much below biologically relevant temperatures).

On the other hand, a similar phenomenon is believed to occur in neutron stars where the temperatures are vastly greater and the condensing species no less massive. Here, the crushing gravity of the neutron stars forces a very high condensate density, pushing up the critical temperature well above the core temperature of the star. But these conditions are obviously not biologically relevant either.

What might be relevant to biology is a very light condensing species, like the Nambu-Goldstone bosons, and this is, in fact, what is postulated in the quantum field theoretical approaches. The condensate density is then a free parameter in terms of which the vacuum phase and the degree of long-range order might be manipulated. The existence of such a parameter is likely to be crucial in determining how quantum and classical mechanisms in the brain might communicate such that, for instance, perceptual and cognitive discriminations might pass from instantiation in networks of neurons to a quantum encryption, and motor action might ultimately be influenced from the quantum subsystem.

Models

The model of Del Giudice *et al.* (1985, 1986) derives from earlier suggestions by Frohlich (1968) that the critical parameter for condensation in biological media might be given in terms of a supply of energy. Indeed, they observe that a polarization is induced in the water surrounding microtubules by even very low amplitude, low frequency electric fields when a supply of energy nourishes the biomolecule. They further note that the required activation energy has approximately the same magnitude as actual metabolic supplies.

The associated spontaneous breakdown of the dipole rotational symmetry yields Nambu-Goldstone modes (dipole wave quanta). Because these modes have such a small mass,

they can condense (actually the transverse optical modes of these quanta mix with photons to produce polaritons, which are the true condensing species) at critical temperatures in a biologically relevant range. Both these authors and Jibu *et al.* (1994) show that this triggers a collective emission mode, known as *superradiance* (Dicke, 1954), that bears some similarity to (but should not be confused with) stimulated emission in a laser. The time scale for long-range correlations that maintain this quantum optical coherence phenomenon is in the range of about 10-14 seconds, much faster than that for thermal processes. Jibu and Yasue (1997) further show that the phenomenon can be protected from thermalization over longer time scales when energy is delivered to the system above a critical pumping rate.

As only one kind of symmetry has been invoked in these models, it may appear that they suffer from an *overprinting* problem; that is, once a vacuum encoding has taken place, no other vacua are accessible for new information without erasing the prior encoding. This was explicitly acknowledged in the papers of Stuart *et al.* (1978, 1979), who suggested that a realistic model would involve a great multiplicity of symmetries. It turns out, however, that overprinting is another problem that evaporates when one moves away from the context of a closed system. In an open, dissipative system like the brain, Vitiello (1995) demonstrates that even a single symmetry leads to a vast number of superposed vacua with a capacity unconstrained by overprinting. Moreover, this opens the way for an understanding of associative memory in terms of "superposition" and "interference" of different vacuum codes and may begin to realize something akin to Pribram's (1971, 1991) suggestions alluded to in the introduction.

As previously mentioned, the condensing species acquires a small, effective mass in a dissipative context. This entails that a supply of energy above a certain threshold will be required, in addition to the external stimulus or "replication signal," to excite the mode into recall.

The spontaneous breakdown of the dipole rotational symmetry leaves a phase symmetry that is in turn spontaneously broken (Del Giudice *et al.*, 1986). In interaction with electromagnetic field, the global phase symmetry becomes a local symmetry, generating an effective mass for the photon by means of the Anderson-Higgs-Kibble mechanism. The resulting, so-called *evanescent* photons are thereby restricted in their range of propagation. Depending on the relative size of the coherence length and the penetration depth, this can lead to *filamentation* of the electric field. The field is restricted to cylindrical tubes due to damping in directions perpendicular to the axis of cylindrical symmetry. This then protects coherence outside the tubes by channeling external fields that are not strong enough to overcome the correlation of the condensate (similar circumstances apply, for instance, in Type II superconductors). Photons impinging on the system that do not have sufficient energy to generate the prescribed effective mass are unable to traverse the system, but their energy might be stored on the network as a polarization mode. When enough energy has been accumulated, the field can escape the region and revert to ordinary (Maxwell) propagation in terms of massless photons. The phases of these photons will, however, be correlated by passage through the system, transferring the coherence of the Nambu-Goldstone modes to the electromagnetic field.

Finally, some suggestions concerning the mechanism by which the quantum and the classical interface in the brain have also been put forward. Jibu and Yasue (1997) have proposed that condensates inside and outside the neural membrane in dendritic nets may

weakly couple, sandwiching the membrane and forming a Josephson junction. Rather than its oft-cited sensitivity to imposed magnetic fields, they draw attention to an electric Josephson effect (Feynman *et al.*, 1965) and demonstrate that the potential difference across the membrane has self-excited oscillations that produce a small, oscillating current across the junction. If many such Josephson junctions span the membrane in an area smaller than the coherence length of the spontaneous symmetry breaking phenomenon, a collective *plasmon* mode emerges and propagates along the membrane as a soliton. Since the soliton wave retains its form, it may propagate over long distances, potentially to macroscopic effect. The Josephson currents across the membrane initiate charge density waves in neurons, suggestive of the depolarizing waves operative at the classical level of neural networks, and perhaps pointing to the link that would allow quantum and classical mechanisms to interface.

References

Quantum Vitalism

Stuart Hameroff

1. What is life?
2. Life mysteries: Protein shape, cell differentiation, and "unitary one-ness"
3. Consciousness and evolution
4. The cytoskeleton, intelligent behavior and differentiation
5. The dawn of consciousness
6. Consciousness and orchestrated objective reduction (Orch OR)
7. Three candidates for Cambrian consciousness
8. Health and disease
9. Conclusion

1. What is life?

Life is a process generally described in terms of its properties and functions including self-organization, metabolism (energy utilization), adaptive behavior, reproduction, and evolution.

Whether or not this functional description is complete is a matter of contention. Two broad types of approaches have historically attempted to characterize the essential nature of the living state:

(1) functionalism and (2) vitalism.

Functionalism implies that life is independent of its material substrate. For example, certain types of self-organizing computer programs exhibit life-like functions, and "artificial life" proponents view such systems as "alive." Functionalists also point out that life's material substrate doesn't distinguish biological matter. Proteins, DNA, carbohydrates, fats, and other components are made of the same mass and elements that make up inanimate substances. Functional /reductive approaches dominate present day molecular biology, demystified by genetic engineering. "Life" is ascribed to an emergent property of biochemical processes and functional activities.

Nonetheless, a commonly held contrary viewpoint is that functional descriptions fail to consider or explain an essential "unitary one-ness" or other property present in living systems. To nineteenth-century biologists this quality was ascribed to a (presumably electromagnetic) "life force," "elan vital," or "life energy ". However as molecular and cell biology began to reveal the biochemical and physical processes involved in cellular activities, the apparent need for a life force waned, and "vitalists" (or "animists") were vilified. Electromagnetic fields associated with cells and tissues were appropriately seen as effects, rather than cause of biological activity. To this day any vitalistic life force or energy field is deemed unnecessary and unacceptable.

Figure 1 Quantum vitalism is an alternative to functionalist/emergent approaches to the problems of both life and consciousness.

The current situation in biology is very much like that of consciousness studies earlier this century. For many years behaviorism dominated psychology and consciousness

(which could not be measured) was, almost literally, taboo. Reductive/functionalist approaches explained the mind as a property of neural information processes. However in the last decade a resurgence of interest in the brain/mind problem has questioned this assumption, symbolized by Chalmers' "hard problem" of the nature of experience, or qualia. Questioning functional explanations of consciousness has led to consideration of relevant macroscopic quantum mechanisms in the brain. If quantum mechanisms participate in consciousness, then quantum mechanisms of some sort must occur more generally throughout biology, presumably preceding the onset of consciousness in the course of evolution. Perhaps it's time to question functionalism and consider quantum vitalism. The basic idea is that life derives by direct extension from dynamics at the fundamental level of reality. Is such a drastic leap necessary? What's wrong with functionalism? As we shall see, mysteries about life persist in the face of functionalist/reductionist science.

Figure 2 The famous experiment by Stanley Miller in which nutrients, atmosphere and electrical spark produced amino acids.

Figure 3. The most widely accepted (though far from proven) idea follows the Miller experiment. Lightning, nutrients and atmosphere produced amino acids which eventually led to proteins, DNA and complex life. From Voet and Voet.

How did life start? Three basic suggestions are that life emerged either from 1) biochemical processes in a warm, wet, soup-like environment (for example as suggested by the famous Miller experiments in which a primordial "soup" of nutrients, oxygen and an electrical spark produced amino acids, Figures 1 and 2), 2) as clay-like hydrated crystals ("mud", suggested by Graham Cairns-Smith), or 3) as "nanoforms"---mineral structures with internal cavities of differing dielectric strength. Whichever of these, or some other, the simple systems somehow supposedly gained abilities to metabolize, adapt and evolve. How? All three examples could have provided a mechanism for a quantum state to be both isolated from environment, and also able in some way to interact with its environment. Proteins begat protein assemblies, and RNA and DNA co-developed (though some argue RNA preceded other biomolecules). Simple organisms, and eventually more complex organisms evolved through differentiation, mutation and selection. How do these occur?

2. Life mysteries: Protein shape, cell differentiation, and "Unitary one-ness"

One mystery of living systems remains the protein folding problem. Despite knowledge of complete sequence and properties of each amino acid in a particular protein, the protein's folding and consequent shape and function cannot be predicted. The problem is described as "NP complete", meaning that in principle it could be solved on a classical

computer, but requiring near infinite time to solve. Some noncomputable feature may be involved. A related mystery is how, once folded, a protein's dynamic conformational state is regulated.

Proteins are made up of a polypeptide string of amino acids, each with a specific amino acid residue (22 possibilities). Let's assume the primordial soup idea is correct. Amino acids were formed, then eventually enough amino acids were around so they began to interact with each other to form polypeptides, and eventually proteins. The early interactions among amino acid residues may be the same as occurs in protein folding: as residues interact, the string folds in three dimensions to determine protein shape. The number of possible interactions among residues and pathways to final shape, or conformation is staggering, unable to be simulated in any reasonable time on the best classical computers. This is the protein folding problem.

Figure 3 Proteins are a string of amino acids whose hydrophobic components (top, above line) self-associate, folding inward to form a hydrophobic pocket.

Figure 4. Amino acids are graded as to their hydrophobicity.

Figure 5 The hydrophobic pocket in the anesthetic-sensitive enzyme papain, shown here with an anesthetic molecule sitting in the pocket.

How does the protein do it? As folding occurs in a liquid environment, the first step is driven by the non-polar hydrophobic residues which are insoluble in water (Figures 3, 4 and 5). They burrow inward, and self-associate, forming a dry hydrophobic pocket sheltered from the aqueous medium. This is entropy driven, as the non-polar groups cannot form hydrogen bonds with water, causing water molecules to order in "clathrates" over the nonpolar surface if it is exposed. When the nonpolar hydrophobic groups bury themselves and stick together in the protein interior forming a pocket, the water ordering is reduced and so entropy drives the system. The process is basically the same as micelle formation with lipids and formation of soaps.

Figure 6. Nonpolar amino acid residues, or anesthetic molecules which are hydrophobic are unstable and relatively insoluble in water. A nonpolar group forces surrounding water to form highly ordered clathrates whose low entropy drives the nonpolar molecule toward a hydrophobic environment. From Voet and Voet.

Figure 7. Clathrate formation From Voet and Voet

So nonpolar groups formed hydrophobic pockets within proteins, sheltered from aqueous environment

So what? Well, long before the first cells, proteins defined an inside and an outside. Proteins sense their environment, and react by changing shape. Proteins are sensitive to various modes of input (temperature, pH, ionic concentration, ligand, phosphorylation etc) and react accordingly by conformational change.

How do proteins process information? Studies from anesthesia suggest the hydrophobic pocket is the brain of the protein. Biologist Robert Rosen characterized an essential feature of life as separation of inside and outside, or self and non-self. Nicholas Humphrey has written in similar terms ("privatization of sensation" as has Maxine Sheets-Johnstone. These have been applied to primitive cells, but proteins and perhaps other biomolecules which preceded cells also have the basic internal-external dichotomy, with internal events interacting with, and responding to environment.

Figure 8 Van der Waals forces are dipole couplings between electron clouds of adjacent atoms or molecular groups. There are permanent dipoles and induced dipoles. London dispersion forces form between adjacent induced dipoles. They are weak, but numerous and influential, and mediate effects of anesthetic gases.

The hydrophobic pockets in certain brain proteins mediate anesthesia, and thus apparently consciousness.

From the study of anesthetic mechanisms we know such pockets can have a solubility and dielectric constant equivalent to that of olive oil, or n-octanol [$\sim 10 \text{ (cal cm}^{-3})^{1/2}$ Halsey, 1974], are about 0.4 cubic nanometers (1/30 to 1/100 of the protein volume) and comprised of nonpolar amino acid residues such as the aromatic tryptophan, with a double resonance ring suitable for electron delocalization. Huddled together, the nonpolar residues form amongst themselves van der Waals London dispersion forces: pairs of

induced dipole electrons in clouds of neighboring residues. The forces are highly dependent on exact distances, falling off as 10^{-6} of the distance between coupled electrons, and very weak (1/100 electron volt), however they are quite numerous (hundreds per pocket) and ultimately influential in mediating protein conformational dynamics. The hydrophobic pocket may be seen as the "quantum brain" of each protein.

Figure 9. London forces in hydrophobic pockets can regulate the protein conformation and function. As quantum mechanical effects they can superposition, leading the protein into superposition according to the Orch OR model.

Another mystery relates to differentiation. All cells in our bodies have a complete set of genes, but each cell is specialized according to which particular subset of genes is expressed. This processing of developing the particular morphological and functional state of a cell or tissue---what makes a muscle cell a muscle cell and a gland a gland---is called differentiation. This specialization apparently came into play early in the course of evolution when simple multicellular colonies of identical independent cells began to communicate and differentiate for the beneficial division of labor. By expressing a particular genetic subset, a specific group of proteins and other biomolecules is synthesized. The selected materials then self-arrange---guided, transported and organized by the cell's internal self-assembling cytoskeleton into a healthy kidney cell, neuron, lymphocyte, or other final product. The cytoskeleton also organizes cell division, forming mitotic spindles and centrioles which separate chromosomes and form daughter cell shape. The state of differentiation is directly related to health.---for example uncontrolled growth and differentiation lead to malignancy/cancer. Cells may also differentiate into "apoptosis" or programmed cell death. Failure to properly maintain differentiation results in "trophic" changes (for example various types of dystrophy, hypertrophy, and a variety of subtle disease states). Conversely, a finely tuned state of differentiation may be optimally healthy.

How is differentiation regulated? Evidence suggests that differentiation in cells and tissues maintained by communicating energy/information from neighboring cells. If cells are removed from their natural habitat in a living organism their state of differentiation and function is lost: they "de-differentiate" to primitive stem cells. To maintain normal differentiation, cells continuously receive coordinated signals ("trophic factors") via molecules secreted by other cells which bind to membrane receptors (for example hormones, cytokines) and also by mechanisms that involve direct communication with adjacent cells. Figure 10 shows three modes of cell-to-cell communication (Clark and Brugge, 1995).

Figure 10 Each set of coupled figures illustrates a different mode of cell-to-cell communication. Left: molecular signal (neurotransmitter, hormone, cytokine etc) secreted by one cell (in this case an axon at top) binds to membrane receptor and initiates response in second cell (bottom). Cytoskeletal structures within "sending" cells enable secretion. Thicker cytoskeletal structures are microtubules interconnected by MAPs (microtubule-associated proteins); thinner cytoskeletal structures are a variety of proteins linking microtubules with membrane proteins and other cell components.

Middle: the region between cells contains the extracellular matrix, a network of filamentous biomolecules that connect to membrane-spanning "integrin" proteins which in turn connect to the cytoskeleton. Right: Two cells connected by "gap junctions" which directly link cell interiors. Cytoplasm, ions and current pass through the gaps between the cells, creating essentially one large cell.

Cells in a tissue are embedded in a fibrous network that fills the space between cells, the "extracellular matrix". The matrix connects directly to each cell's outer surface via a class of proteins called integrins which span the membrane and communicate with cytoskeletal networks in the cell interior. The cytoskeleton is a network of filamentous proteins including microtubules, actin, and intermediate filaments which determine cell shape and coordinate cellular activities. Direct mechanical signaling along the extracellular matrix, integrins, and cytoskeleton has been demonstrated and appears to regulate gene expression and differentiation (Bissell, Hall and Parry, 1982; Puck and Krystosek, 1992; Maniotis, Bojanowski and Ingber, 1997; Maniotis, Chen and Ingber, 1997).

Many types of cells are also interconnected by direct focal connections. Among these are "gap junctions", membrane protein channels between adjacent cells through which cytoplasm, ions and current flow. Neurons interconnected by gap junctions form networks that "fire synchronously, behaving like one giant neuron" (Kandel et al, 1991). Gap junctions are essential in embryology, coordinating groups of cells during organ development and also enable communication in mature organs (for example liver cells are regulated by gap junctions).

In both types of direct connections, cells yield individual autonomy and become integrated into tissues and organs displaying "unitary oneness". The integration appears at least partially mediated by information represented by dynamical conformational states of membrane proteins, cytoskeleton, and extracellular matrix. Exactly how individual proteins regulate their conformational states is unknown, however as we have heard quantum interactions in hydrophobic pocket may be the key.

If quantum states do regulate individual proteins and other biomolecules, what happened as biomolecules developed into larger structures---assemblies of proteins like microtubules, organelles, cells, tissues, organs, complex organisms? Did the quantum states remain at the level of individual proteins, as Michael Conrad would suggest? Or did the quantum states also grow, and become increasingly macroscopic in protein assemblies, organelles, cells, tissues, organs and organisms?

Mechanisms for isolation from environmental decoherence would be needed, and could have co-evolved (Lecture week 7). Gap junction connections between cells seem a potentially convenient mode of inter-cellular spread of quantum states, although as Scott Hagan has pointed out quantum field theoretical approaches predict quantum spread across membranes as well. If quantum states do indeed exist within biomolecules and

cells and play the role of information arbiter in proteins, assemblies and so forth, they would seem obligatory in the eventual evolution of consciousness.

3. Consciousness and Evolution

(The following material is from "Did consciousness cause the Cambrian evolutionary explosion, from the MIT Press Tucson II book and on my website)

When and where did consciousness emerge in the course of evolution? Did it happen as recently as the past million years, for example concomitant with language or tool making in humans or primates? Or did consciousness arrive somewhat earlier, with the advent of mammalian neocortex 200 million years ago (Eccles, 1992)? At the other extreme, is primitive consciousness a property of even simple unicellular organisms of several billion years ago (e.g. as suggested by Margulis and Sagan, 1995)? Or did consciousness appear at some intermediate point, and if so, where and why? Whenever it first occurred did consciousness alter the course of evolution?

According to fossil records, life on earth originated about 4 billion years ago (Figure 1).

/

~

Figure 11. The Cambrian explosion. According to fossil records life on earth originated about 4 billion years ago, but evolved only slowly for about 3.5 billion years ("pre-Cambrian period"). Then, in a rather brief 10 million years beginning about 540 million years ago (the "Cambrian period"), a vast array of diversified life abruptly emerged: the "Cambrian explosion." Exemplary Cambrian organisms depicted are an urchin similar to present day actinospira, spiny worms, and a tentacled sectorian. Artwork by Dave Cantrell and Cindi Laukes based on organisms in Gould (1989) and adapted from a diagram by Joe Lertola, Time Magazine, December 4, 1995.

For its first 3.5 billion years or so ("pre-Cambrian period") life seems to have evolved slowly, producing only single cells and a few simple multicellular organisms. The most significant life forms for the 2 billion years of this period were algae and bacteria-like prokaryotes. Then, about 1.5 billion years ago, eukaryotic cells appeared apparently as symbiotic mergers of previously independent organ (mitochondria, plastids) with prokaryotic cells. According to biologist Lynn Margulis (Margulis, 1975; Margulis and Sagan, 1995) microtubules and the dynamically functional cytoskeleton were also of additions, originating as independent motile spirochetes which invaded prokaryotes and formed a mutually favorable symbiosis. Prokaryotic cells provided a stable, nourishing environment, biochemical energy to the spirochetes who reciprocated by cytoskeletal-based locomotion, sensation, mitosis and differentiation. Pre-Cambrian eukaryotic cells continued to slowly evolve for another billion or so years, resulting only in simple multicellular organisms. Then, in a rather brief 10 million years beginning about 540 million years ago (the beginning of the "Cambrian period"), there apparently occurred a world-wide dramatic acceleration in the rate of evolution: the "Cambrian explosion." A vast array of diversified life abruptly emerged: all the phyla from which today's animals are descended (e.g. Gould, 1989). The Cambrian explosion theory has been questioned. For example, using fossil nucleotide substitution analysis, Wray et al. (1996) suggested

a more linear process, with animals appearing about one billion years ago. But the more gradual, linear case assumes constant rate of nucleotide substitution. It seems more likely that nucleotide substitution also increases during increased rates of evolution, and the "abrupt" Cambrian explosion theory still holds (Vermeij, 1996). What could have precipitated the Cambrian explosion? Were climate, atmosphere, environment or some external factors important, or did a threshold of biological genetic complexity occur (e.g. Kauffman, 1995; c.f. Dawkins, 1989)? Can a particular biological functionality be identified that critically enhanced adaptation, survivability and mutation? Did purposeful, intelligent behavior accelerate evolution? The idea that behavior can directly alter genetic code formed the basis of an eighteenth century evolutionary theory by Jean-Baptiste Lamarck. No supportive evidence was found to show that behavior directly modified genetics, and "Lamarckism" was appropriately discredited. The question of whether behavior can alter the course of evolution indirectly was discussed by Schrödinger (1958) who offered several examples (Scott, 1996). A species facing predators and harsh environment might best survive by producing a large number of offspring for cooperative support. Such a rapidly reproducing species is ripe for accelerated evolutionary development (Margulis and Sagan 1995). Another example is a species which finds a new habitat (moving onto land, climbing trees. . .) to which adaptation is facilitated by supporting mutations. Changes in behavior can also favor chance mutations which reinforce original changes, resulting in closed causal loops or positive feedback in evolutionary development (Scott, 1996). Generally, intelligent behavior can enhance a species' survivability and the opportunity for mutation by avoiding extinction. How did intelligent behavior come to be? Dennett (1995) describes the "birth of agency": the ability to perform purposeful actions in complex macromolecules, and thus very early in the course of evolution. He emphasizes that agency and behavior at the macromolecular level are non-conscious and clearly preceded Cambrian multicellular organisms. For example purposeful behavior surely occurred in unicellular eukaryotic ancestors of modern organisms like paramecia and euglena who perform rather complex adaptive movements. Paramecia swim in a graceful, gliding fashion via coordinated actions of hundreds of microtubule-based cilia on their outer surface. They thus seek and find food, avoid obstacles and predators, and identify and couple with mates to exchange genetic material. Some studies suggest paramecia can learn, escaping more quickly from capillary tubes with each subsequent attempt (Gelber, 1958). Having no synapses or neural networks, paramecia and similar organisms rely on their cytoskeleton for sensation, locomotion and information processing. The cytoskeleton organizes intelligent behavior in eukaryotic cells.

4. The Cytoskeleton, Intelligent Behavior and Differentiation

Comprising internal scaffolding and external appendages of each eukaryotic cell, the cytoskeleton includes microtubules (MTs), actin filaments, intermediate filaments and complex arrays of connected MTs such as centrioles, cilia, flagella and axonemes. MTs are hollow cylinders 25 nanometers (nm) in diameter (Figure 12).

Figure 12. Left: Microtubule (MT) structure: a hollow tube of 25 nanometers diameter, consisting of 13 columns of tubulin dimers arranged in hexagonal lattice (Penrose, 1994). Right (top): Each tubulin molecule can switch between two (or more) conformations, coupled to a quantum even such as electron location in tubulin hydrophobic pocket. Right (bottom): Each tubulin can also exist in quantum superposition of both conformational states (Hameroff and Penrose, 1996).

Their lengths vary and may be quite long within some nerve processes. MT cylinder walls are hexagonal lattices of tubulin subunit proteins polar, 8 nm peanut-shaped dimers which consist of two slightly different 4 nm monomers (alpha and beta tubulin). MTs are interlinked by a variety of MT-associated-proteins to form dynamic networks which define cell shape and functions. Numerous types of studies link the cytoskeleton to cognitive processes (for review, cf. Dayhoff et. al., 1994; Hameroff and Penrose, 1996a). Theoretical models and simulations suggest that conformational states of tubulins within MT lattices are influenced by quantum events, and can interact with neighboring tubulins to represent, propagate and process information in classical "cellular automata," or ferroelectric "spin-glass" type computing systems (e.g. Hameroff and Watt, 1982; Rasmussen et al., 1990; Tuszynski et al, 1995). There is some suggestion that quantum coherence could be involved in MT computation (Jibu et al., 1995). MTs are often assembled in nine MT doublets or triplets in a mega-cylinder found in centrioles, cilia and flagella (Figure 13).

~

Figure 13. Primitive appendages comprised of microtubules (MTs, shown as circles; scale bars: 100 nanometers). Top left: Cross section of a centriole. Nine microtubule triplets link to a single central MT. Top right: Cilium in cross section. Nine MT doublets link to a central MT pair. Bottom left (A): cross section of a suctorian tentacle in an open, dilated phase. Bottom right (B): suctorian tentacle in a constricted phase. (Tentacle structure adapted from Hauser and Van Eys, 1976, by Dave Cantrell.) Scale bar: 100 nanometers.

Centrioles are two MT mega-cylinders in perpendicular array which control cell cycles and mitosis, form the focal point of the rest of the cell cytoskeleton, and provide cell navigation and orientation. Embedded in dense electronegative material just adjacent to the cell nucleus, centrioles' structural beauty, unfathomable geometry and intricate behavior have created an air of mystery: "biologists have long been haunted by the possibility that the primary significance of centrioles has escaped them" (Wheatley, 1982; c.f. Lange and Gull, 1996).

Cilia are membrane covered mega-cylinders of nine MT doublets with an additional central MT pair which are both sensory and motor: cilia can receive information, as well as move in a coordinated fashion for locomotion, feeding or movement of material. Flagella have the same 9&2 MT arrangement as cilia, but are longer and more specialized for rapid cell movement. The basic microtubule doublet "9&2" (cilia, flagella, basal bodies...) and "9&0"(centriole) arrangements apparently originated in spirochetes

prior to eukaryotes. Their cytoskeletal descendants provided agency to eukaryotes, performing a variety of purposeful behaviors.

Cytoskeletal structures also provided internal organization, variation in cell shape, separation of chromosomes and differentiation. An essential factor in evolution, differentiation involves emergence of specialized tissues and organs from groups of cells which start out alike, but began to differ and develop specific and complementary form and functions (Rasmussen, 1996). Each eukaryotic cell contains all the genes of the organism, but only a subset are selected so that, for example, liver cells are distinct from lymphocytes. Puck (e.g. Puck and Krystosek, 1992) has shown appealing evidence to suggest that a cell's genes are activated and regulated by its cytoskeleton, and describes how differentiation requires cooperative cytoskeletal function. Tissue specialization also required factors such as actin-gelation phases in cytoplasm, MT-based extensions (cilia, axonemes. . .) and communication among cells. The most basic and primitive form of inter-cellular communication involves direct cell-cell channels such as gap junctions, or electrotonic synapses (Lo, 1985; Llinas, 1985). Cytoskeletal cooperativity among neighboring cells enabled differentiation and allowed different types of tissues to emerge. Through benefit of the resultant division of labor, higher order structures (organs e.g. axonemes, tentacles, eye cups, nervous systems. . .) with novel functions appeared (Rasmussen, 1996). These in turn led to more intelligent behavior in small multicellular animals.

5. The Dawn of Consciousness

According to this scenario, tissue differentiation, agency and intelligent behavior were occurring for a billion years from the symbiotic origin of eukaryotes to the Cambrian explosion (Figure 1). What then happened? Was some critical level of intelligent behavior suddenly reached? Did consciousness then appear? Could primitive consciousness have significantly improved fitness and survivability beyond previous benefit provided by non-conscious agency and intelligent behavior?

One possible advantage of consciousness for natural selection is the ability to make choices. As Margulis and Sagan (1995) observe (echoing similar, earlier thoughts by Erwin Schrödinger), "If we grant our ancestors even a tiny fraction of the free will, consciousness, and culture we humans experience, the increase in [life's] complexity on Earth over the last several thousand million years becomes easier to explain: life is the product not only of blind physical forces but also of selection in the sense that organisms choose. . ." (Scott, 1996).

By itself, the ability to make choices is insufficient evidence for consciousness (e.g. computers can choose intelligently). However non-computable, seemingly random conscious choices with an element of unpredictability may have been particularly advantageous for survival in predator-prey dynamics (e.g. Barinaga, 1996).

Another feature of consciousness favoring natural selection could be the nature of conscious experience: qualia, our "inner life" in the sense of Chalmers' "hard problem" (e.g. Chalmers, 1996a; 1996b). Organisms which are not conscious, but have intelligent behavior are (in the philosophical sense) "zombies." If a zombie organism is threatened but has no experience of fear or pain, it may not react decisively. A conscious organism having an experience of fear or pain would be motivated to avoid threatening situations, and one having experience of taste would be more motivated to find food. The experience of pleasure could well have promoted reproductive efforts.

Who were the early Cambrian organisms? Fossil records have identified a myriad of small worms, strange urchins, tiny shellfish and many other creatures (Gould, 1989) depicted in the bottom of Figure 1. Nervous systems among small Cambrian worms (by comparison with apparent present-day cousins like the nematode worm *C. elegans*) may be estimated to contain roughly hundreds of neurons. Primitive eye cups and vision were also prevalent, as were tube-like alimentary systems with a mouth at one end and anus at the other. Cambrian urchins and other creatures also featured prominent spine-like extensions seemingly comparable to axoneme spines in present day echinoderms such as *actinosphaerium*. The versatile axonemes (MT arrays more complex than those of cilia and centrioles) are utilized for sensation, locomotion and manipulation, and provide perception, agency and purposeful, intelligent behavior.

As consciousness can't be measured or observed in the best of circumstances, it seems impossible to know whether or not consciousness emerged in early Cambrian organisms (or at any other point in evolution). The simple (hundreds of neuron) neural networks, primitive vision, purposeful spine-like appendages and other adaptive structures which characterize early Cambrian creatures depend heavily on cytoskeletal function and suggest the capability for agency, intelligent behavior and the possibility of primitive consciousness. Perhaps coincidentally, a specific model (Orch OR) predicts the occurrence of consciousness at precisely this level of cytoskeletal size and complexity.

6. Consciousness and orchestrated objective reduction (Orch OR)

What is conscious experience? Believing contemporary understanding of brain function inadequate to explain "qualia," or experience, a line of panpsychist/panexperiential philosophers (e.g. Leibniz, Whitehead, Wheeler, Chalmers. . .) have concluded that consciousness derives from an experiential medium which exists as a fundamental feature of reality. If so, conscious experience may be in the realm of spacetime physics, and raw "protoconscious" information may be encoded in spacetime geometry at the fundamental Planck scale [e.g. Penrose's (1971) quantum spin networks-Rovelli and Smolin, 1995a; 1995b]. A self-organizing Planck-scale quantum process could select "fundamental" experience resulting in consciousness. Is there such a process?

A self-organizing quantum process operating at the interface between quantum and macroscopic states, objective reduction (OR) is Penrose's (1989; 1994; 1996) quantum gravity solution to the problem of wave function collapse in quantum mechanics. According to quantum theory (and repeatedly verified experimentally), small scale quantum systems described by a wave function may be "superposed" in different states and/or places simultaneously. Large scale "macroscopic systems," however, always appear in definite "classical" states and/or places. The problem is that there is no apparent reason for this "collapse of the wave function," no obvious border between microscopic quantum and macroscopic classical conditions. The conventional explanation (the "Copenhagen interpretation") is that measurement or observation by a conscious observer collapses the wave function. To illustrate the apparent absurdity of this notion, Schrodinger (1935) described a now-famous thought experiment in which a cat is placed in a box into which poison is released when triggered by a particular quantum event. Schrodinger pointed out that according to the Copenhagen interpretation, the cat would be both dead and alive until the box was opened and the cat observed by a conscious human.

To explain this conundrum, many physicists now believe that intermediate between tiny quantum-scale systems and "large" cat-size systems some objective factor disturbs the superposition causing collapse, or reduction to classical, definite states and locations. This putative process is called objective reduction (OR). One increasingly popular OR viewpoint (initiated by Karolyhazy in 1966 Karolyhazy, et al., 1986) suggests this "largeness" is to be gauged in terms of gravitational effects and in Einstein's general relativity, gravity is spacetime curvature. According to Penrose (1989; 1994; 1996), quantum superposition actual separation (displacement) of mass from itself causes underlying spacetime to also separate at the Planck scale due to simultaneous curvatures in opposite directions. Such separations are unstable and a critical degree of separation (related to quantum gravity) results in spontaneous self-collapse (OR) to particular states chosen non-computably.

In Penrose's OR the size of an isolated superposed system (gravitational self-energy E of a separated mass) is inversely related to the coherence time T according to the uncertainty principle $E = h/T$, where h (actually " $\hbar$ ") is Planck's constant over 2π . T is the duration of time for which the mass must be superposed to reach quantum gravity threshold for self-collapse. Large systems (e.g. Schrodinger's 1 kg cat) would self-collapse (OR) very quickly, in only 10^{-37} seconds. An isolated superposed single atom would not OR for 10^6 years. Somewhere between those extremes are brain events in the range of tens to hundreds of milliseconds. A 25 millisecond brain event (i.e. occurring in coherent 40 Hz oscillations) would require nanogram (10^{-9} gram) amounts of superposed neural mass.

In the Penrose-Hameroff "Orch OR" model (e.g. Penrose and Hameroff, 1995; Hameroff and Penrose, 1996a; 1996b), quantum coherent superposition develops in microtubule subunit proteins ("tubulins") within brain neurons and glia. The quantum state is isolated from environmental decoherence by cycles of actin gelation, and connected among neural and glial cells by quantum tunneling across gap junctions (Hameroff, 1996). When the quantum gravity threshold is reached according to $E = h/T$, self-collapse (objective reduction) abruptly occurs. The pre-reduction, coherent superposition ("quantum computing") phase is equated with pre-conscious processes, and each instantaneous OR, or self-collapse, corresponds with a discrete conscious event. Sequences of events give rise to a "stream" of consciousness. Microtubule-associated-proteins "tune" the quantum oscillations and the OR is thus self-organized, or "orchestrated" ("Orch OR"). Each Orch OR event selects microtubule subunit states non-computably which classically regulate synaptic/neural functions. Because the superposed protein mass separation is also a separation in underlying spacetime geometry, each Orch OR event selects a particular "fundamental" experience.

Consider a low intensity conscious sensory perception, for example as activity in sensory cortex after lightly touching a finger. Such an event was shown by Libet et al (1979) to have a pre-conscious time of 500 msec until conscious awareness. For $T = 500$ msec of quantum coherent superposition, $E =$ self-collapse of approximately 10^9 tubulins. As typical neurons contain about 10^7 tubulins, Orch OR predicts involvement of roughly 10^2 to 10^3 neurons (interconnected by gap junctions) for rudimentary conscious events. For more intense conscious events, for example consistent with 25 msec "cognitive quanta" defined by coherent 40 Hz activity (e.g. Crick and Koch, 1990; c.f.

Llinas, e.g. Joliot et al, 1994), superposition and self-collapse of 2×10^{10} tubulins (and 10^3 to 10^4 neurons) would be required.

How might Orch OR have happened? One possibility is that quantum coherence emerged in eukaryotic MT assemblies via the Fröhlich mechanism as a by-product of coordinated dynamics and biochemical energy (e.g. Frohlich, 1968; 1970; 1975). Quantum coherence could also be an intrinsic property of the structure and geometry of microtubules and centrioles introduced to eukaryotes by spirochetes. Development of actin gels provided isolation for MT quantum states, and inter-cellular gap junction connections (suitable for quantum tunneling) enabled larger and larger quantum states among MTs in many connected cells. At some point in the course of evolution, sufficient quantum coherence to elicit Orch OR by $E = h/T$ was reached. Rudimentary "conscious" events then occurred. Organisms began to have experience and make conscious, non-computable choices.

7. Three candidates for Cambrian consciousness

Here three biological scenarios consistent with the Orch OR model for early Cambrian emergence of consciousness are considered. Each case involves MTs containing a minimum of 10^9 tubulins, suitable for a 500 msec conscious event.

1) sufficiently complex gap junction-connected neural networks (hundreds of neurons e.g. small worms).

2) primitive vision (ciliated ectoderm eye cup-e.g. small worms).

3) geometrical microtubule arrays (e.g. axoneme spines in small urchins such as actinosphaerium, tentacles in suctorians).

Many early Cambrian fossils are small worms with simple nervous systems. In Hameroff and Penrose (1996b) we speculated that among current organisms, the threshold for rudimentary Orch OR conscious events (500 msec pre-conscious time) may be very roughly at the level of 300 neuron (3×10^9 neural tubulin) nematode worms such as the well studied *C. elegans*. This should be roughly the same neural network complexity as early Cambrian worms which could apparently burrow, swim, or walk the ocean floor with tentacles and spines (Figure 11; Gould, 1989).

Another candidate for the Cambrian emergence of Orch OR consciousness involves the evolution of visual photoreceptors. Amoeba respond to light by diffuse sol-gel alteration of their actin cytoskeleton (Cronly-Dillon and Gregory, 1991). Euglena and other single cell organisms have localized "eye spots" e.g. regions at the root of the microtubule-based flagellum. Cytoplasm may focus incident light toward the eye spots and pigment material shields certain angles to provide directional light detection (e.g. Insinna, this volume). Euglena swim either toward or away from light by flagellar motion. Having no neurons or synapses, the single cell euglena's photic response (sensory, perceptive and motor components) depend on MT-cytoskeletal structures.

Mammalian cells including our own can respond to light. Albrecht-Buehler (e.g. 1994) showed that single fibroblast cells move toward red/infra-red light by utilizing their MT-based centrioles for directional detection and guidance ("cellular vision"); he also points out that centrioles are ideally designed photodetectors (Figure 14).

Figure 14. Photoreception/phototransduction mechanisms at all stages of evolution involve the nine MT doublet or triplet structures found in centrioles, cilia, flagella and axonemes. Left: The centriole is a pair of MT-based mega-cylinders arrayed in perpendicular (Lange and Gull, 1996). Albrecht-Buehler (1994) has identified centrioles as the photoreceptor/phototransducer in photosensitive eukaryotic cells. Middle: Flagellar axonemes are the photosensitive structures in protozoa such as Euglena gracilis. Right: Cilia in rod and cone retinal cells in vertebrate eyes (including humans) bridge two parts of the cells. Photosensitive pigments (rhodopsin) is contained in the outer segment (top) while cell nucleus, mitochondria and synaptic connection are contained in the cell body (bottom). Light enters the eye (from the bottom in this illustration) and traverses the cell body and cilium to reach the rhodopsin-containing outer segment. Adapted from Lange and Gull (1996) and Insinna (1997) by Dave Cantrell. Scale bars: 100 nanometers.

~

Figure 15. A schematic representation of the process of superradiance in a microtubule proposed by Mari Jibu, Kunio Yasue and colleagues (Jibu et al., 1994). Each oval without an arrow stands for a water molecule in the lowest rotational energy state. Each oval with an arrow stands for a water molecule in the first excited rotational state. The process is cyclic (a to b to c to d to a b) and so on. A) Initial state of the system of water molecules in a microtubule. Energy gain due to the thermal fluctuations of tubulins increases the number of water molecules in the first excited rotational energy state. B) A collective mode of the system of water molecules in rotationally excited states. A long-range coherence is achieved inside a microtubule by means of spontaneous symmetry breaking. C) A collective mode of the system of water molecules in rotationally excited states loses its energy collectively, and creates coherent photons in the quantized electromagnetic field inside a microtubule. D) Water molecules, having lost their first excited rotational energies by superradiance, start again to gain energy from the thermal fluctuation of tubulins, and the system of water molecules recover the initial state (a).

With permission from Jibu et al. (1994).

Jibu et al (1994; 1996) have predicted that cellular vision depends on a quantum state of ordered water in MT inner cores (Figure 15).

They postulate a nonlinear quantum optical effect termed "superradiance" conveying evanescent photons by a process of "self-induced transparency" (the optical analogue of superconductivity). Hagan (1995) has observed that cellular vision provided an evolutionary advantage for single cell organisms with cilia, centrioles or flagella capable of quantum coherence.

In simple multicellular organisms, eyes and visual systems began with groups of differentiated light-sensitive ciliated cells which formed primitive "eye cups" (up to 100 photoreceptor cells) in many phyla including flatworms, annelid worms, molluscs, crustacea, echinoderms and chordates (our original evolutionary branch Cronly-Dillon and Gregory, 1991). The retinas in our eyes today include over 108 rod and cone photoreceptors each comprised of an inner and outer segment connected by a ciliated

stalk. As each cilium is comprised of about 300,000 tubulins, our retinas contain about 3×10^{13} tubulins per eye. (Retinal rods, cones and glia are interconnected by gap junctions- Leibovic, 1990.) Conventional vision science assumes the cilium is purely structural, but the centriole/cilium/flagella MT structure which Albrecht-Buehler has analyzed as an ideal directional photoreceptor may detect or guide photons in eye spots of single cells, primitive eye cups in early multicellular organisms, and rods and cones in our retinas. Quantum coherence leading to consciousness could have emerged in sheets of gap junction-connected ciliated cells in eye cups of early Cambrian worms. Perhaps consciousness occurred in even simpler organisms? Many Cambrian fossils are similar or related to present day species having particularly interesting MT geometrical arrangements. For example actinosphaerium (echinosphaerium) nucleofilum is a present-day echinoderm, a tiny sea-urchin heliozoan with about one hundred rigid protruding axonemes about 300 microns in length (Figure 4). Appearing similar to spines of Cambrian echinoderms, actinosphaerium axonemes sense and interact with environment, provide locomotion, and are each comprised of several hundred MTs interlinked in a double spiral (Figure 16).

~

Figure 16. Cross-section of double spiral array of interconnected MTs in single axoneme of actinosphaerium, a tiny heliozoan related to sea urchin echinoderms present at the Cambrian evolutionary explosion (Figure 1). Each cell has about one hundred long and rigid axonemes which are about 300 microns long, made up of a total of 3×10^9 molecules of tubulin (Roth et al., 1970; Dustin, 1985). Scale bar: 500 nm (with permission from L.E. Roth).

Each axoneme contains about 3×10^7 tubulins, and the entire heliozoan contains 3×10^9 tubulins (Roth et al., 1970) perhaps coincidentally the precise quantity predicted by Orch OR for a 500 msec conscious event.

Allison and Nunn (1968; c.f. Allison et al., 1970) studied living actinosphaerium in the presence of the anesthetic gas halothane. They observed that the axoneme MTs disassembled in the presence of halothane (although at anesthetic concentrations two to four times higher than that required for anesthetic effect).

Somewhat similar to axonemes are larger prehensile tentacles in suctorians such as akinetoposis and heliophyra. Small multicellular animals, suctorians (Figure 37.1) have dozens of tiny hollow tentacles which probe their environment and capture and ingest prey such as paramecium. The prehensile tentacles range from about 300 microns up to one millimeter in length. Their internal structure is comprised of longitudinal arrays of about 150 microtubules in a ring around an inner gullet through which prey/food passes (Figure 37.3). MTs apparently slide over one another in a coordinated fashion to provide tentacle movement and contractile waves involved in food capture, ingestion and other adaptive behaviors. The activity is interrupted by the anesthetic halothane (Hauser and Van Eys, 1976). A single suctorian tentacle (150 MTs, length 500 microns) contains about 10^9 tubulins the predicted requirement for a 500 msec Orch OR event. Perhaps consciousness arose in the probings of a Cambrian suctorian tentacle?

Would such primitive Orch OR experiences in a Cambrian worm, urchin or suctorian be anything like ours? What would it be like to be a tentacle? A single, 109 tubulin, 500 msec Orch OR in a primitive system would have gravitational self-energy (and thus experiential intensity) perhaps equivalent to a "touch lightly on the finger" experience. However our everyday coherent 40 Hz brain activity would correspond to 25 msec events involving 2×10^{10} tubulins, and so our typical experience would be some 20 times more intense. We also would have many more Orch OR events per second (e.g. 40 vs maximum of 2) with extensive sensory processing and associative memory presumably lacking in Cambrian creatures. Nonetheless, by Orch OR criteria, a 109 tubulin, 500 msec Orch OR event in a Cambrian worm, urchin or tentacle would be a conscious experience: a smudge of awareness, a shuffle in funda-mental spacetime.

To conclude, the place of consciousness in evolution is unknown, but the actual course of evolution itself may offer a clue. Fossil records indicate that animal species as we know them today including conscious humans all arose from a burst of evolutionary activity some 540 million years ago (the "Cambrian explosion"). It is suggested here that:

- 1) Occurrence of consciousness was likely to have accelerated the course of evolution.
- 2) Small worms, urchins and comparable creatures reached critical biological complexity for emergence of primitive consciousness at the early Cambrian period 540 million years ago.
- 3) Cooperative dynamics of microtubules, cilia, centrioles and axonemes were the critical biological factors for consciousness.
- 4) Cytoskeletal complexity available in early Cambrian animals closely matches criteria for the Penrose-Hameroff Orch OR model of consciousness.

Orch OR caused the Cambrian explosion.

8. Health and disease

Quantum vitalism has implications for medicine and health. In Eastern medicine and increasingly in integrative Western medicine, the body's health or vitality at tissue, organ, and holistic levels is described (somewhat vaguely) in terms of "energy" or "information." In oriental medicine, including acupuncture, the energy is known as "ch'i" (or "Qi," or one of a number of slightly different variants), and its characteristics are appreciated in somewhat greater detail. In Indian medicine and philosophy the energy is known as "prana", or "kundalini". For example, Eastern medicine indicates the energy/information is focused in epicenters called "chakras," and flows along meridians that run through the body but do not precisely follow the somatic nervous system nor obvious anatomical boundaries (although meridians and chakras do correspond somewhat to the autonomic nervous system, and embryological planes). The energy seems to have an electromagnetic correlate or component; however, efforts to fully characterize its physical basis have been unsuccessful.

Perhaps ch'i, or prana is related to the quantum fields? Cells within tissues are interconnected frequently by gap junctions. Ebb and flow of sol-gel transitions within cells, perhaps coordinated among cells, would elicit and extinguish quantum states along meridians, for example.

Dr Andy Weil who is the champion of integrative medicine which attempts to synthesize the best of modern medicine and alternative medicine of various sorts has an interesting observation about the possible relevance of quantum mechanisms to health and disease. Laboratory tests and x-rays, Dr Weil explains, are like measurements on a quantum

system. The tissue may be in superposition of both health and disease with at least the possibility of complete health until the measurement is made. If disease is found, there it is. He is suggesting that some times it may be better not to do the tests, and to treat possible underlying causes and predispositions. The idea of NOT doing lab tests is of course anathema to modern medicine (highly reductionist) and could be construed as malpractice in some contexts. I don't endorse it personally but it is something interesting to think about.

9. Conclusion

Is life subtly connected to a basic level of reality?

From conclusions drawn about consciousness I'd have to say that life involves quantum states in biomolecular systems. They may come and go, ebb and flow, spatiotemporally in cytoplasm, among cells, through tissues along certain embryologically determined meridian pathways. In fact the edge between quantum states and classical states may be close to life itself. Decoherence by environmental interaction may not be random if the cytoplasmic environment isn't random. (which it isn't). Edward Burian has suggested that collapse or decoherence of any kind has the essential feature of consciousness. Maintaining that consciousness requires Penrose OR, I'd amend Burian's suggestion slightly and suggest that decoherence/collapse in biological systems may be at least the edge of life. The unitary one-ness of living systems requires quantum coherence or entanglement. Long live quantum vitalism.